



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Evasão no Bacharelado em Ciência da Computação da Universidade de Brasília: análise e mineração de dados

Luísa Behrens Palmeira
Matheus Parreiras Santos

Monografia apresentada como requisito parcial
para conclusão do Bacharelado em Ciência da Computação

Orientadora
Prof.^a Dr.^a Maria Emilia Machado Telles Walter

Coorientador
Prof. Dr. Jan Mendonça Corrêa

Brasília
2014

Universidade de Brasília — UnB
Instituto de Ciências Exatas
Departamento de Ciência da Computação
Bacharelado em Ciência da Computação

Coordenador: Prof. Dr. Homero Luiz Piccolo

Banca examinadora composta por:

Prof.^a Dr.^a Maria Emilia Machado Telles Walter (Orientadora) — CIC/UnB
Prof. Dr. Jan Mendonça Corrêa — CIC/UnB
Dr.^a Maria Inez Machado Telles Walter — DPO/UnB

CIP — Catalogação Internacional na Publicação

Palmeira, Luísa Behrens.

Evasão no Bacharelado em Ciência da Computação da Universidade de Brasília: análise e mineração de dados / Luísa Behrens Palmeira, Matheus Parreiras Santos. Brasília : UnB, 2014.

287 p. : il. ; 29,5 cm.

Monografia (Graduação) — Universidade de Brasília, Brasília, 2014.

1. evasão, 2. mineração de dados, 3. análise estatística, 4. unb,
5. ciência da computação

CDU 004.4

Endereço: Universidade de Brasília
Campus Universitário Darcy Ribeiro — Asa Norte
CEP 70910-900
Brasília-DF — Brasil



Evasão no Bacharelado em Ciência da Computação da Universidade de Brasília: análise e mineração de dados

Monografia apresentada como requisito parcial
para conclusão do Bacharelado em Ciência da Computação

Prof.^a Dr.^a Maria Emilia Machado Telles Walter (Orientadora)
CIC/UnB

Prof. Dr. Jan Mendonça Corrêa Dr.^a Maria Inez Machado Telles Walter
CIC/UnB DPO/UnB

Prof. Dr. Homero Luiz Piccolo
Coordenador do Bacharelado em Ciência da Computação

Brasília, 06 de junho de 2014

Dedicatória

Dedicamos esse trabalho às nossas famílias, Mônica Behrens Azevedo Palmeira, Luiz Alberto de Almeida Palmeira e Camilla Behrens Palmeira, pais e irmã de Luísa Palmeira, Siumara Maria Parreiras Santos, Último Gonçalves dos Santos e Marcella Parreiras Santos, pais e irmã de Matheus Parreiras, por todo o suporte, paciência e carinho ao longo dos anos de nossa formação.

Agradecimentos

Agradecemos, primeiramente, a Deus, aos nossos pais, irmãs e toda família que, com muito carinho e apoio, não mediram esforços para que chagássemos até essa etapa de nossas vidas. Um agradecimento muito especial também aos nossos amigos que nos deram força para seguir nessa jornada.

Agradecemos à nossa orientadora, Prof.^a Dr.^a Maria Emilia Machado Telles Walter, que aceitou o desafio de orientar um projeto já em andamento, sempre estando ao nosso lado e nos auxiliando. Somos gratos pela consideração, paciência, convívio, apoio e compreensão dedicados a esse trabalho. Sua orientação foi essencial para o desenvolvimento desse projeto, além de ser um exemplo de pessoa e profissional dedicada para nos espelharmos.

Agradecemos ao Prof. Dr. Jan Mendonça Corrêa por ter aceitado o desafio de ser nosso coorientador e pela sua participação, nos apoiando e transmitindo seus conhecimentos de mineração de dados, essenciais para o desenvolvimento desse projeto. Somos extremamente agradecidos à Dr.^a Maria Inez Machado Telles Walter pela cooperação com o nosso trabalho, sabemos que ele não seria possível sem sua atenção e ajuda ao compartilhar os dados necessários para esse trabalho.

Agradecemos imensamente ao nosso amigo Lucas Lo Ami que nos ajudou muito nesse projeto, nos guiando e auxiliando em tudo que precisávamos. Agradecemos pela paciência, pelo carinho, pelo interesse e pela participação fundamental em determinadas etapas do trabalho, além da amizade que foi de suma importância para nos mantermos motivados durante todo o processo.

Agradecemos à Empresa Júnior de Computação - CJR e ao Movimento Empresa Júnior (MEJ) pela maturidade adquirida e por fazer parte de nossa formação, nos oferecendo o primeiro contato com o mercado de trabalho. Sem dúvida umas das melhores experiências que tivemos, nos deixando recordações, conhecimentos e muitas amizades especiais.

Um agradecimento especial do autor Matheus Parreiras Santos à Universidad del Pais Vasco pela grande oportunidade, através do Ciência Sem Fronteiras, de ter a experiência de estudar e viver um intercâmbio. Ao meus professores e amigos estrangeiros e brasileiros, principalmente aos que viveram e estudaram comigo, agradeço pelo carinho que tiveram ao longo dessa maravilhosa experiência.

Resumo

Esse trabalho tem como objetivo analisar os fatores que levam à evasão dentro do Bacharelado em Ciência da Computação da Universidade de Brasília e identificar o perfil dos alunos do curso que estão em risco de evasão. Para isso, são utilizados dados pessoais e do histórico do aluno. A evasão é um problema de nível mundial nas universidades. A quantidade de alunos que ingressam em um curso ou instituição e saem sem se formar gera prejuízos em diversos âmbitos. No curso analisado, a taxa de evasão é de mais de 50%. Buscando determinar as causas do problema para reduzir esse valor, é feita uma análise estatística para identificar os principais fatores relacionados com o perfil de um aluno evadido para, então, aplicar técnicas de mineração de dados que permitem verificar quais os fatores que possuem maior impacto e qual o perfil dos alunos que estão cursando atualmente. Foram desenvolvidos diferentes classificadores visando encontrar o que possui melhor performance para os dados analisados. Os resultados mostram que o problema da evasão está se agravando no curso.

Palavras-chave: evasão, mineração de dados, análise estatística, unb, ciência da computação

Abstract

This work aims to analyze the factors that lead to dropout rate in the Computer Science Bachelor degree from the University of Brasilia and identify the students' profiles that are at risk of dropping. Personal and transcript data are used. Dropout rate is a global problem in universities. The number of students enrolled in a course or institution that leave without graduating generates losses in different areas. Over the Computer Science Bachelor degree, the dropout rate is higher than 50 %. Aiming to determine the causes of the problem to reduce this value, statistical analysis is performed to identify the main factors related to the profile of an dropout student to then apply data mining techniques to check what factors have the greatest impact and the profile of the students who are currently attending. Different classifiers were developed to meet the highest performing to the analyzed data. The results show that the dropout rate problem is getting worse in the course.

Keywords: dropout rate, data mining, statistical analysis, unb, computer science

Sumário

1	Introdução	1
1.1	Problema	1
1.2	Hipótese	1
1.3	Objetivos	2
1.4	Organização do trabalho	2
2	Educação Superior, Evasão e o BCC – UnB	3
2.1	Educação Superior no Brasil	3
2.2	Conceito de Evasão	6
2.3	Estudos sobre evasão	6
2.4	Bacharelado em Ciência da Computação - UnB	8
2.4.1	Informações sobre o curso	8
2.4.2	Disciplinas obrigatórias	9
2.4.3	Disciplinas optativas	10
3	Mineração de Dados	14
3.1	Apresentação	14
3.2	Conceitos básicos	14
3.2.1	Mineração de dados e Extração de conhecimento	15
3.2.2	Objetivos da extração de conhecimento	17
3.2.3	Arquitetura de um sistema de mineração de dados	17
3.3	Tipos de aprendizado	19
3.3.1	Aprendizado supervisionado	19
3.3.2	Aprendizado não supervisionado	19
3.3.3	Aprendizado por reforço	19
3.4	Técnicas de mineração de dados	20
3.4.1	Classificação	20
3.4.2	Regressão	22
3.4.3	Clusterização	22
3.5	Ferramentas de mineração de dados	23
4	Trabalhos Relacionados	26
4.1	Visão geral	26
4.2	Evasão na UnB	27
4.3	Evasão no Brasil	28
4.4	Comparações dos trabalhos	31
4.4.1	Dados	32

4.4.2	Metodologias	32
5	Análise de Dados no BCC - UnB	37
5.1	Evasão para esse trabalho	37
5.2	Aquisição dos dados	38
5.3	Tratamento dos dados	41
5.4	Perguntas definidas	43
5.5	Análise de dados	44
5.5.1	Semestre	44
5.5.2	Sexo	45
5.5.3	Idade	46
5.5.4	Tipo de escola	47
5.5.5	Forma de ingresso	47
5.5.6	Desempenho médio	49
5.5.7	Taxa de reprovação média	50
5.5.8	Média de créditos obrigatórios por semestre	50
5.5.9	Índice de reprovação por departamento	51
5.5.10	Reprovações em disciplinas obrigatórias	51
5.5.11	Taxa de reprovação ao longo dos anos	52
5.5.12	Menções em disciplinas obrigatórias	54
6	Perfil dos Alunos do BCC – UnB	69
6.1	Transformação dos dados	69
6.2	Mineração dos dados	70
6.3	Avaliação dos algoritmos	73
6.4	Alunos com perfil de evasão	76
7	Conclusões	78
7.1	Principais contribuições	78
7.2	Trabalhos futuros	79
	Referências	80
I	Ementas e pré-requisitos das disciplinas obrigatórias	83
A	Consulta SQL para tratamento dos dados	91
B	Consulta SQL para análise estatística dos dados	95
C	Dados da análise estatística	99
D	Consulta SQL para mineração de dados	115

Lista de Figuras

3.1	Etapas do processo de extração de conhecimento. Adaptado de Han e Kamber [19].	16
3.2	Arquitetura típica de um sistema de mineração de dados. Adaptado de Han e Kamber [19].	18
3.3	Exemplo da etapa de treinamento da classificação, segundo Han e Kamber [19]. Nesse exemplo, a classificação terá objetivo de determinar se um empréstimo é de risco ou é seguro. Os arquivos de entrada já possuem exemplos de algumas pessoas e, com base nesses dados, o algoritmo de classificação determina as regras da classificação.	20
3.4	Exemplo da segunda etapa da classificação, segundo Han e Kamber [19]. Esse exemplo representa a segunda etapa do que foi representado na Figura 3.3. Os arquivos de testes, que já estão classificados mas não foram utilizados na fase de treinamento, rodam sobre as regras de classificação para verificar a acurácia do classificador. Caso a acurácia esteja satisfatória para a aplicação, nossos dados iram rodar sobre essas regras de classificação para determinar se o empréstimo é de risco ou não.	21
3.5	Interface <i>Explorer</i> do Weka versão 3.6 no Windows.	24
3.6	Interface <i>Knowledge Flow</i> do Weka versão 3.6 no Windows.	24
3.7	Interface <i>Experimenter</i> do Weka versão 3.6 no Windows.	25
5.1	Taxa de evasão por semestre e motivos de evasão em cada um. O código do motivo de saída está de acordo com o apresentado na Tabela 5.3. O semestre 0 corresponde a um semestre cursado no verão, independente da posição do fluxo em que o aluno estava quando cursou. Cada uma das barras corresponde a um semestre, representado no eixo X. A altura da barra representa a quantidade de alunos que evadiram naquele semestre, de acordo com o eixo Y. As cores em cada uma das barras correspondem ao motivo de saída, segundo a legenda do lado direito do gráfico. O número dentro de cada uma das cores mostra a quantidade absoluta de alunos que evadiram naquele semestre e por qual motivo.	45
5.2	Taxa de formatura por semestre. Cada uma das barras corresponde a um semestre, representado no eixo X. A altura da barra representa a quantidade de alunos que evadiram naquele semestre, de acordo com o eixo Y.	46

5.3	Taxa de evasão e de formatura feminina e motivos de saída de quem evadiu. O código do motivo de saída está de acordo com o apresentado na Tabela 5.3. O gráfico do lado esquerdo mostra a quantidade de evadidos, de formados e de alunos cursando, sendo o azul a quantidade de alunos cursando, o vermelho a de formados e o laranja a de evadidos. Já o do lado direito, mostra os motivos de saída dos evadidos, sendo os números no gráfico a quantidade de alunos que saíram por cada motivo.	47
5.4	Taxa de evasão e de formatura masculina e motivos de saída de quem evadiu. O código do motivo de saída está de acordo com o apresentado na Tabela 5.3. O gráfico do lado esquerdo mostra a quantidade de evadidos, de formados e de alunos cursando, sendo o azul a quantidade de alunos cursando, o vermelho a de formados e o laranja a de evadidos. Já o do lado direito, mostra os motivos de saída dos evadidos, sendo os números no gráfico a quantidade de alunos que saíram por cada motivo.	48
5.5	Índice de evasão por idade e por motivo de saída. O código do motivo de saída está de acordo com o apresentado na Tabela 5.3. Cada uma das barras corresponde a uma idade, representada no eixo X. A altura da barra representa a quantidade de alunos que evadiram com aquela idade, de acordo com o eixo Y. As cores em cada uma das barras correspondem ao motivo de saída, segundo a legenda do lado direito do gráfico. O número dentro de cada uma das cores mostra a quantidade absoluta de alunos que evadiram com aquela idade e por esse motivo.	50
5.6	Índice de formatura por idade. Cada uma das barras corresponde a uma idade, representada no eixo X. A altura da barra representa a quantidade de alunos que evadiram com aquela idade, de acordo com o eixo Y.	51
5.7	Taxa de evasão por tipo de escola e motivos de saída. O código do motivo de saída está de acordo com o apresentado na Tabela 5.3. Cada uma das barras corresponde a um tipo de escola, representado no eixo X. A altura da barra representa a quantidade de alunos que evadiram que possuem aquele tipo de escola, de acordo com o eixo Y. As cores em cada uma das barras correspondem ao motivo de saída, segundo a legenda do lado direito do gráfico. O número dentro de cada uma das cores mostra a quantidade absoluta de alunos que evadiram daquele tipo de escola e por esse motivo. .	52
5.8	Taxa de evasão por forma de ingresso. O código da forma de ingresso está de acordo com o apresentado na Tabela 5.6. Cada uma das barras corresponde a uma forma de ingresso, representada no eixo X. A altura da barra representa a quantidade de alunos que evadiram que possuem aquele tipo de forma de ingresso, de acordo com o eixo Y. As barras vermelhas correspondem aos alunos que evadiram e a verde ao total que entraram por aquela forma de ingresso, de acordo com a legenda do lado direito do gráfico.	53
5.9	Desempenho médio dos evadidos organizados em faixas de desempenho. Cada uma das barras corresponde a uma faixa de desempenho, representada no eixo X. A altura da barra representa a quantidade de alunos evadidos que possuem desempenho dentro daquela faixa, de acordo com o eixo Y.	53

5.10	Desempenho médio dos formados organizados em faixas de desempenho. Cada uma das barras corresponde a uma faixa de desempenho, representada no eixo X. A altura da barra representa a quantidade de alunos formados que possuem desempenho dentro daquela faixa, de acordo com o eixo Y.	54
5.11	Taxa de reprovação média dos evadidos organizados em faixas de reprovação. Cada uma das barras corresponde a uma faixa de taxa de reprovação, representada no eixo X. A altura da barra representa a quantidade de alunos evadidos que possuem taxa de reprovação dentro daquela faixa, de acordo com o eixo Y.	54
5.12	Taxa de reprovação média dos formados organizados em faixas de reprovação. Cada uma das barras corresponde a uma faixa de taxa de reprovação, representada no eixo X. A altura da barra representa a quantidade de alunos formados que possuem taxa de reprovação dentro daquela faixa, de acordo com o eixo Y.	54
5.13	Média de créditos obrigatórios cursados por semestre dos evadidos. O semestre 0 representa um semestre cursado no verão, independente do período curricular que ele estava quando cursou. Cada uma das barras corresponde a um semestre, representada no eixo X. A altura da barra representa a média de créditos obrigatórios cursados por semestre, de acordo com o eixo Y.	55
5.14	Média de créditos obrigatórios cursados por semestre dos formados. O semestre 0 representa um semestre cursado no verão, independente do período curricular que ele estava quando cursou. Cada uma das barras corresponde a um semestre, representada no eixo X. A altura da barra representa a média de créditos obrigatórios cursados por semestre, de acordo com o eixo Y.	55
5.15	Índice de reprovação das matérias obrigatórias organizadas por departamento dos alunos evadidos e dos formados. Os departamentos estão representados no eixo X. A altura da barra representa a taxa de reprovação deles, independente da situação do estudante no curso, de acordo com o eixo Y.	56
5.16	Reprovações nas disciplinas obrigatórias recomendadas de serem cursadas no 1º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10. Cada uma das barras corresponde a uma disciplina, representada no eixo X. A altura da barra representa a quantidade de reprovações que já ocorreram naquela disciplina, independente da situação do estudante no curso, de acordo com o eixo Y.	56
5.17	Reprovações nas disciplinas obrigatórias recomendadas de serem cursadas no 2º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10. Cada uma das barras corresponde a uma disciplina, representada no eixo X. A altura da barra representa a quantidade de reprovações que já ocorreram naquela disciplina, independente da situação do estudante no curso, de acordo com o eixo Y.	57

5.18	Reprovações nas disciplinas obrigatórias recomendadas de serem cursadas no 3º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10. Cada uma das barras corresponde a uma disciplina, representada no eixo X. A altura da barra representa a quantidade de reprovações que já ocorreram naquela disciplina, independente da situação do estudante no curso, de acordo com o eixo Y. .	57
5.19	Reprovações nas disciplinas obrigatórias recomendadas de serem cursadas no 4º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10. Cada uma das barras corresponde a uma disciplina, representada no eixo X. A altura da barra representa a quantidade de reprovações que já ocorreram naquela disciplina, independente da situação do estudante no curso, de acordo com o eixo Y. .	58
5.20	Reprovações nas disciplinas obrigatórias recomendadas de serem cursadas no 5º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10. Cada uma das barras corresponde a uma disciplina, representada no eixo X. A altura da barra representa a quantidade de reprovações que já ocorreram naquela disciplina, independente da situação do estudante no curso, de acordo com o eixo Y. .	58
5.21	Evolução da taxa de reprovação ao longo dos anos nas disciplinas obrigatórias recomendadas de serem cursadas no 1º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10. Cada linha está relacionada a uma disciplina, de acordo com a legenda do lado direito do gráfico. Cada um dos pontos dessas linhas relaciona a taxa de reprovação em um determinado ano. A taxa de reprovação foi calculada como a proporção de reprovações das matrículas em uma determinada disciplina a cada ano.	59
5.22	Evolução da taxa de reprovação ao longo dos anos nas disciplinas obrigatórias recomendadas de serem cursadas no 2º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10. Cada linha está relacionada a uma disciplina, de acordo com a legenda do lado direito do gráfico. Cada um dos pontos dessas linhas relaciona a taxa de reprovação em um determinado ano. A taxa de reprovação foi calculada como a proporção de reprovações das matrículas em uma determinada disciplina a cada ano.	60
5.23	Evolução da taxa de reprovação ao longo dos anos nas disciplinas obrigatórias recomendadas de serem cursadas no 3º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10. Cada linha está relacionada a uma disciplina, de acordo com a legenda do lado direito do gráfico. Cada um dos pontos dessas linhas relaciona a taxa de reprovação em um determinado ano. A taxa de reprovação foi calculada como a proporção de reprovações das matrículas em uma determinada disciplina a cada ano.	61

5.24	Evolução da taxa de reprovação ao longo dos anos nas disciplinas obrigatórias recomendadas de serem cursadas no 4º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10. Cada linha está relacionada a uma disciplina, de acordo com a legenda do lado direito do gráfico. Cada um dos pontos dessas linhas relaciona a taxa de reprovação em um determinado ano. A taxa de reprovação foi calculada como a proporção de reprovações das matrículas em uma determinada disciplina a cada ano.	62
5.25	Evolução da taxa de reprovação ao longo dos anos nas disciplinas obrigatórias recomendadas de serem cursadas no 5º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10. Cada linha está relacionada a uma disciplina, de acordo com a legenda do lado direito do gráfico. Cada um dos pontos dessas linhas relaciona a taxa de reprovação em um determinado ano. A taxa de reprovação foi calculada como a proporção de reprovações das matrículas em uma determinada disciplina a cada ano.	63
5.26	Menções nas disciplinas obrigatórias recomendadas de serem cursadas no 1º semestre do curso dos alunos formados e dos evadidos. Cada uma das barras corresponde a uma menção, representada no eixo X. A altura da barra representa a proporção de cada uma das menções obtidas naquela disciplina, de acordo com o eixo Y.	64
5.27	Menções nas disciplinas obrigatórias recomendadas de serem cursadas no 2º semestre do curso dos alunos formados e dos evadidos. Cada uma das barras corresponde a uma menção, representada no eixo X. A altura da barra representa a proporção de cada uma das menções obtidas naquela disciplina, de acordo com o eixo Y.	65
5.28	Menções nas disciplinas obrigatórias recomendadas de serem cursadas no 3º semestre do curso dos alunos formados e dos evadidos. Cada uma das barras corresponde a uma menção, representada no eixo X. A altura da barra representa a proporção de cada uma das menções obtidas naquela disciplina, de acordo com o eixo Y.	66
5.29	Menções nas disciplinas obrigatórias recomendadas de serem cursadas no 4º semestre do curso dos alunos formados e dos evadidos. Cada uma das barras corresponde a uma menção, representada no eixo X. A altura da barra representa a proporção de cada uma das menções obtidas naquela disciplina, de acordo com o eixo Y.	67
5.30	Menções nas disciplinas obrigatórias recomendadas de serem cursadas no 5º semestre do curso dos alunos formados e dos evadidos. Cada uma das barras corresponde a uma menção, representada no eixo X. A altura da barra representa a proporção de cada uma das menções obtidas naquela disciplina, de acordo com o eixo Y.	68

6.1	Exemplo de arquivo no formato ARFF. O nome da relação entre os atributos é definido por @relation. Cada um dos atributos e seus possíveis valores são declarados por @attribute. O início da listagem dos dados é identificado por @data. Cada registro está em uma linha e os valores dos atributos são separados por vírgula e estão na mesma ordem da declaração dos atributos.	72
6.2	Conversão de valores numéricos para nominais utilizando o filtro do Weka. Os números estão de acordo com os passos apresentados.	72
6.3	<i>Confusion matrix</i> gerada pelo classificador construído com o algoritmo <i>Naive Bayes</i> . Essa matriz mostra quantos alunos, dos que estão cursando, são classificados como formandos e quantos como evadidos . O número total de instâncias é 263.	77
6.4	Proporção de alunos cursando que foram classificados como formandos ou evadidos pelo classificador construído com o algoritmo <i>Naive Bayes</i> .	77

Lista de Tabelas

2.1	Cenário do ensino superior em 2012 com base em estatísticas da educação brasileira realizada pelo INEP [12].	4
2.2	Quantidade de matrículas e cursos por área de conhecimento em 2012 com base em estatísticas da educação brasileira realizada pelo Inep [12], sendo N a quantidade total de cada tipo de curso e matrícula.	4
2.3	Quantidade de ingressantes e concluintes em cursos de graduação presenciais do ensino superior em 2012 com base em estatísticas da educação brasileira realizada pelo Inep [12].	5
2.4	Quantidade de ingressantes e concluintes em cursos de graduação presenciais de Ciências da Computação do ensino superior em 2012 com em estatísticas da educação brasileira realizada pelo Inep [12].	5
2.5	Disciplinas obrigatórias do BCC – UnB [10].	9
2.6	Cadeia 4: Ciclo 3: o aluno deverá cursar [10].	10
2.7	Cadeia 5: Ciclo 3: o aluno deverá cursar [10].	10
2.8	Cadeia 6: Ciclo 3: o aluno deverá cursar [10].	10
2.9	Cadeia 7: Ciclo 3: o aluno deverá cursar [10].	10
2.10	Cadeia 8: Ciclo 3: o aluno deverá cursar [10].	10
2.11	Disciplinas optativas do Bacharelado em Ciência da Computação da UnB [10].	11
4.1	Tabela com as descrições e finalidade de cada indicador que compõe o método de regressão logística, segundo Vitelli et al. [34].	30
4.2	Métodos de mineração de dados e respectivas descrição e finalidades, segundo Campello e Lins [6].	31
4.3	Categorias de dados que foram utilizados nos métodos estatísticos dos trabalhos analisados por esse projeto.	33
4.4	Objetivos apresentados nos trabalhos analisados que compuseram o entendimento do fenômeno de evasão estudado.	33
4.5	Formas de realização da extração dos dados apresentadas nos trabalhos analisados que compuseram o entendimento do fenômeno de evasão estudado.	34
4.6	Métodos estatísticos apresentados nos trabalhos analisados que compuseram o entendimento do fenômeno de evasão estudado.	34
4.7	Períodos analisados que compuseram o entendimento do fenômeno de evasão estudado.	35
4.8	Dados que foram descartados apresentados nos trabalhos analisados que compuseram o entendimento do fenômeno de evasão estudado.	35
4.9	Procedimentos de tratamento de dados apresentados nos trabalhos analisados que compuseram o entendimento do fenômeno de evasão estudado. .	36

4.10	Novas variáveis criadas nos trabalhos analisados que compuseram o entendimento do fenômeno de evasão estudado.	36
5.1	Variáveis relacionadas aos estudantes do Bacharelado em Ciência da Computação.	38
5.2	Variáveis relacionadas ao histórico dos estudantes já matriculados no Bacharelado em Ciência da Computação.	40
5.3	Código e descrição das possíveis formas de saída de um aluno.	42
5.4	Fórmulas utilizadas para cálculo do semestre curricular em que uma disciplina foi cursada em função do semestre em que ela foi feita e do que o aluno entrou no curso. A coluna semestre_entrada corresponde ao semestre de entrada do aluno em algum ano e a semestre_cursada corresponde ao semestre de algum ano em que a disciplina foi cursada. Nas fórmulas, a variável ano_entrada é o ano de entrada do aluno no curso e ano_cursada é o ano em que a disciplina foi cursada.	43
5.5	Quantidade de alunos que já saíram do curso separados por sexo e situação em que eles saíram.	46
5.6	Código e descrição das possíveis formas de ingresso de um aluno.	48
6.1	Variáveis escolhidas para serem utilizadas no processo de mineração de dados. A escolha das variáveis foi feita de acordo com os fatores que possuem maior impacto na definição do perfil, com base nos resultados obtidos na análise estatística dos dados, apresentada no Capítulo 5.	70
6.2	Nome dos algoritmos utilizados para classificação nesse trabalho e o nome da classe deles na ferramenta Weka.	73
6.3	Performance dos classificadores criados com base em critérios gerais do classificador.	74
6.4	Performance dos classificadores criados com base em critérios específicos da classe formados	74
6.5	Performance dos classificadores criados com base em critérios específicos da classe evadidos	74
6.6	Performance dos classificadores criados com base em critérios gerais do classificador.	75
6.7	Performance dos classificadores após o teste com novos dados com base em critérios específicos da classe formados	75
6.8	Performance dos classificadores após o teste com novos dados com base em critérios específicos da classe evadidos	76
I.1	Ementas e pré-requisitos das disciplinas obrigatórias do BCC – UnB [10]. .	83
C.1	Índice de evasão por semestre e por motivo de saída. O código do motivo de saída está de acordo com o apresentado na Tabela 5.3. O semestre 0 corresponde a um semestre cursado no verão, independente da posição do fluxo em que o aluno estava quando cursou.	99
C.2	Índice de formatura por semestre. O semestre 0 corresponde a um semestre cursado no verão, independente da posição do fluxo em que o aluno estava quando cursou.	101

C.3	Taxa de evasão feminina e motivos de saída. O código do motivo de saída está de acordo com o apresentado na Tabela 5.3.	102
C.4	Taxa de evasão masculina e motivos de saída. O código do motivo de saída está de acordo com o apresentado na Tabela 5.3.	102
C.5	Índice de evasão por idade e por motivo de saída. O código do motivo de saída está de acordo com o apresentado na Tabela 5.3.	102
C.6	Índice de formatura por idade.	105
C.7	Taxa de evasão por tipo de escola e motivos de saída. O código do motivo de saída está de acordo com o apresentado na Tabela 5.3.	106
C.8	Taxa de evasão por forma de ingresso. O código da forma de ingresso está de acordo com o apresentado na Tabela 5.6. A quantidade de evadidos representa quantos alunos que entraram por cada uma das formas de ingresso e evadiram. O total de alunos é a quantidade de alunos que entraram por aquela forma de ingresso, independente da situação atual dele no curso, ou seja, se está cursando, formado ou evadido.	107
C.9	Desempenho médio dos evadidos organizados em faixas de desempenho. . .	107
C.10	Desempenho médio dos formados organizados em faixas de desempenho. . .	107
C.11	Taxa de reprovação média dos evadidos organizados em faixas de reprovação.	108
C.12	Taxa de reprovação média dos formados organizados em faixas de reprovação.	108
C.13	Média de créditos obrigatórios cursados por semestre dos evadidos. O semestre 0 representa um semestre cursado no verão, independente do período curricular que ele estava quando cursou.	108
C.14	Média de créditos obrigatórios cursados por semestre dos formados. O semestre 0 representa um semestre cursado no verão, independente do período curricular que ele estava quando cursou.	109
C.15	Índice de reprovação das matérias obrigatórias organizadas por departamento dos alunos evadidos e dos formados.	109
C.16	Reprovações nas disciplinas obrigatórias recomendadas de serem cursadas no 1º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10.	110
C.17	Reprovações nas disciplinas obrigatórias recomendadas de serem cursadas no 2º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10.	110
C.18	Reprovações nas disciplinas obrigatórias recomendadas de serem cursadas no 3º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10.	110
C.19	Reprovações nas disciplinas obrigatórias recomendadas de serem cursadas no 4º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10.	110
C.20	Reprovações nas disciplinas obrigatórias recomendadas de serem cursadas no 5º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10.	111
C.21	Evolução da taxa de reprovação ao longo dos anos nas disciplinas obrigatórias recomendadas de serem cursadas no 1º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10.	111

C.22	Evolução da taxa de reprovação ao longo dos anos nas disciplinas obrigatórias recomendadas de serem cursadas no 2º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10.	112
C.23	Evolução da taxa de reprovação ao longo dos anos nas disciplinas obrigatórias recomendadas de serem cursadas no 3º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10.	112
C.24	Evolução da taxa de reprovação ao longo dos anos nas disciplinas obrigatórias recomendadas de serem cursadas no 4º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10.	113
C.25	Evolução da taxa de reprovação ao longo dos anos nas disciplinas obrigatórias recomendadas de serem cursadas no 4º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10.	113
C.26	Menções obtidas pelos alunos evadidos nas disciplinas obrigatórias. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10.	114
C.27	Menções obtidas pelos alunos formados nas disciplinas obrigatórias. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10.	114

Capítulo 1

Introdução

O ensino superior, não apenas no Brasil mas em todos os países do mundo, possui uma importância muito grande tanto para o estudante quanto para a sociedade, pois, como citado por Ferreira et al. [17], ele pode ser considerado como um dos fatores fundamentais para o desenvolvimento econômico, social e cultural nas nações, em um contexto de crescente competição global.

Um grande problema que atinge o ensino superior atualmente, e que não se restringe ao Brasil, é a quantidade de alunos que ingressam em uma instituição de ensino e, por algum motivo, saem dela sem se formar. Essa questão da evasão gera consequências graves em diversos âmbitos, tanto do sistema educacional quanto aos próprios estudantes, numa dimensão social e financeira.

Por conta disso, tratar a evasão e a retenção de alunos tem sido um dos desafios do ensino superior, como explicado por Sarker et al. [29].

Nesse contexto, tendo em vista a falta de estudos sobre evasão no Bacharelado em Ciência da Computação (BCC – UnB), não se tem uma visão clara do desempenho do curso atualmente.

Os resultados de análise de informações das causas, disciplinas ou algum conjunto de fatores sobre a evasão no BCC – UnB e o delineamento de um perfil de um aluno com risco de evasão pode auxiliar a Coordenação de Graduação do curso em dar suporte específico a esses alunos e a orientar todos os demais alunos.

1.1 Problema

Nesse contexto, o problema desse trabalho é que não existem estudos sobre evasão no BCC – UnB, ofertado pelo Departamento de Ciência da Computação da UnB.

1.2 Hipótese

A fim de solucionar o problema desse trabalho, essa pesquisa trabalhará com a hipótese de que é possível determinar o perfil de um aluno em risco de evasão no curso de Bacharelado em Ciência da Computação da UnB utilizando técnicas de mineração de dados.

1.3 Objetivos

Objetivo geral

Analisar os fatores que levam à evasão dentro do BCC – UnB e identificar o perfil de um aluno em risco de evasão usando uma abordagem baseada em mineração de dados.

Objetivos específicos

- Realizar análises estatísticas dos dados de evasão do curso de Bacharelado em Ciência da Computação, a partir de dados do SIGRA.
- Propor técnica de mineração de dados para analisar os dados de evasão.
- Utilizar a técnica no Weka.
- Analisar os dados de evasão do BCC (obtidos do SIGRA).
- Discutir os resultados das estatísticas e da mineração de dados.

1.4 Organização do trabalho

O presente trabalho está organizado nos seguintes capítulos. O Capítulo 2 apresenta uma visão geral da situação do ensino superior e dos conceitos básicos relacionados com a evasão e trata sobre o BCC – UnB que será o estudo de caso desse trabalho. O Capítulo 3 explica os conceitos e o processo de mineração de dados, além das ferramentas utilizadas nesse trabalho. O Capítulo 4 apresenta alguns trabalhos já realizados sobre estudo de evasão utilizando técnicas de análise e mineração de dados. O Capítulo 5 mostra o resultado da análise estatística dos dados e o Capítulo 6 apresenta o processo realizado utilizando mineração de dados. As conclusões são apresentadas no Capítulo 7.

Capítulo 2

Educação Superior, Evasão e o BCC – UnB

Nesse capítulo, inicialmente será discutido sobre Educação Superior e Evasão e, em seguida, será descrito o PPP do BCC – UnB.

Na Seção 2.1 é apresentado o cenário atual da educação superior no Brasil e na Seção 2.2 é definido o termo evasão. A Seção 2.3 apresenta alguns estudos sobre evasão, que nesse trabalho foram analisados com o objetivo de auxiliar no entendimento do contexto atual e na definição desse problema. A Seção 2.4 apresenta informações sobre o BCC – UnB e suas disciplinas optativas e obrigatórias.

2.1 Educação Superior no Brasil

Uma forma conveniente de estudar o cenário da educação superior do Brasil é analisando os dados fornecidos pelo INEP ¹ (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira), órgão do Ministério da Educação ².

O INEP tem divulgado com frequência e de forma padronizada dados estatísticos da educação superior em seu portal de dados ³, que possuem informações como dados gerais das instituições de ensino, matrículas, concluintes em cursos e processos seletivos.

De acordo com os dados divulgados pelo INEP em 2012, percebe-se, na Tabela 2.1, que ocorre uma predominância de Instituições de Ensino Superior (IES) privadas no Brasil, correspondendo 87.4% das IES brasileiras. Vale ressaltar que dentro de cada tipo de IES, ou seja, pública e privada, estão sendo consideradas as diferentes formas de organização acadêmica, como, universidades, centros universitários, faculdades, IF e CELET.

Ainda segundo os dados da Tabela 2.1, observa-se a predominância das IES privadas tanto na quantidade de cursos oferecidos (65.8%) quanto na quantidade de alunos matriculados (73%).

Já a Tabela 2.2 mostra os dados da quantidade de cursos e de alunos matriculados por área de conhecimento de acordo com o tipo de IES (pública ou privada). É interessante notar que, com exceção de cursos da área de Agricultura e Veterinária, a quantidade de

¹<http://www.inep.gov.br>

²<http://www.mec.gov.br/>

³<http://portal.inep.gov.br/>

Tabela 2.1: Cenário do ensino superior em 2012 com base em estatísticas da educação brasileira realizada pelo INEP [12].

Tipo de IES	IES		Cursos		Matricula	
	Quantidade	%	Quantidade	%	Quantidade	%
Pública	304	12.6	10.905	34.2	1.897.374	27.0
Privada	2.112	87.4	20.961	65.8	5.140.312	73.0
Total	2.416	100	31.866	100	7.037.688	100

alunos matriculados em cursos de cada área em IES privadas é superior a de alunos dessa mesma área matriculados em instituições públicas.

Tabela 2.2: Quantidade de matrículas e cursos por área de conhecimento em 2012 com base em estatísticas da educação brasileira realizada pelo Inep [12], sendo N a quantidade total de cada tipo de curso e matrícula.

Área de conhecimento	Pública				Privada			
	Cursos		Matrículas		Cursos		Matrículas	
	N	%	N	%	N	%	N	%
Educação	4.471	41.1	601.167	31.8	3.700	17.6	761.068	14.8
Humanidades e Artes	618	5.2	61.281	3.3	836	4.0	98.726	1.9
Ciências Sociais, Negócios e Direito	1.698	15.6	424.606	22.6	7.854	37.4	2.472.257	48.1
Ciências, Matemática e Computação	1.143	10.5	170.434	9.1	2.033	9.7	260.580	5.1
Engenharia, Produção e Construção	1.336	12.3	268.542	14.3	2.431	11.6	617.370	12.0
Agricultura e Veterinária	553	5.1	102.311	5.6	327	1.6	62.764	1.2
Saúde e Bem-Estar Social	846	7.9	215.492	11.5	2.861	13.7	745.831	14.5
Serviços	240	2.3	31.362	1.8	919	4.4	121.365	2.4
Total	10.905	100	1.897.376	100	20.961	100	5.140.312	100

As estatísticas da Educação Superior realizadas pelo INEP fornecem diversos outros dados que permitem analisar a qualidade da educação no Brasil. Desde a criação da Lei nº10.861, de 14 de abril de 2004, o Sistema Nacional de Avaliação da Educação Superior (SINAES) é utilizado para avaliar a qualidade da educação superior brasileira. O Sinaes avalia aspectos como ensino, pesquisa, responsabilidade social, desempenho de alunos e infraestrutura. Para atingir seu objetivo, o SINAES utiliza instrumentos complementares: auto-avaliação, avaliação externa, ENADE ⁴, avaliação de cursos de graduação e informações de censos e cadastros. Os processos avaliativos são coordenados e supervisionados pela Comissão Nacional de Avaliação da Educação Superior (CONAES) e a operacionalização é de responsabilidade do INEP.

Galvão et al. [22] questionam se o sistema de avaliação do SINAES é o único modelo capaz de avaliar a qualidade do ensino superior no Brasil ou se outras propostas poderiam

⁴<http://portal.inep.gov.br/enade>

trazer contribuições mais proveitosas. Ainda nesse estudo, eles propuseram um modelo de avaliação global para instituições de ensino superior. Bertolin [5] também propôs um sistema de indicadores para avaliar o desenvolvimento e a qualidade da educação superior brasileira em nível de sistema, baseando-se em modelos já existentes no Brasil e no mundo. Dois pontos que ambos colocaram como indicador e que vêm sendo estudados em todo o mundo devido à importância para a educação são a evasão e a retenção.

O conceito de evasão será abordado com mais profundidade na subseção adiante, mas, para uma breve análise da situação brasileira desses indicadores de evasão e retenção, é importante definir como eles podem ser vistos.

Um estudo preliminar da situação da educação brasileira em relação a esses indicadores pode ser realizado por meio da observação, na Tabela 2.3, da quantidade de estudantes que ingressam em uma instituição de ensino superior e da quantidade daqueles que concluem seus estudos no mesmo período. Nessa relação, o esperado seria que os dois dados fossem semelhantes, já que, enquanto uma nova turma está ingressando, outra turma de mesma quantidade de alunos deveria estar se formando. Analisando os dados provenientes do censo do INEP, é possível ter uma noção da situação brasileira acerca dos indicadores de evasão e de retenção e verificar que a realidade está longe do que é esperado.

Tabela 2.3: Quantidade de ingressantes e concluintes em cursos de graduação presenciais do ensino superior em 2012 com base em estatísticas da educação brasileira realizada pelo Inep [12].

Tipo de IES	Ingressantes	Concluintes	Porcentagem
Pública	1.715.752	202.394	11.8%
Privada	4.208.086	673.697	16.0%
Total	5.923.838	876.091	14.8%

Como pode ser visto, apenas 14.8% da quantidade de matriculados concluem seus estudos no mesmo período, caracterizando os 85.2% restantes como retidos ou evadidos no ensino superior. Também pode ser notado que as instituições públicas possuem essa diferença mais acentuada do que as privadas. Nas públicas, apenas 11.8% concluem seus estudos contra 16% nas instituições particulares. Analisando os números dos cursos de Ciências da Computação, apresentados na Tabela 2.4, nota-se a mesma situação da esfera nacional, mas com uma queda nos índices dos cursos fornecidos por instituições públicas. Os cursos de Ciências da Computação possuem apenas 14.3% concluintes da quantidade de matriculados naquele período, mas apresentam uma queda para 9.4% do valor comparando com as instituições públicas.

Tabela 2.4: Quantidade de ingressantes e concluintes em cursos de graduação presenciais de Ciências da Computação do ensino superior em 2012 com em estatísticas da educação brasileira realizada pelo Inep [12].

Tipo de IES	Ingressantes	Concluintes	Porcentagem
Pública	43.681	4.088	9.4%
Privada	115.280	18.700	16.2%
Total	158.961	22.788	14.3%

Esses números demonstram a importância dada por Bertolin [5] na inclusão dos indicadores de evasão e de retenção para estudo do ensino superior no país.

2.2 Conceito de Evasão

Existem diversas definições para o termo evasão nos estudos sobre o assunto no ensino superior. Santos et al. [15] consideram como evasão a situação em que um aluno que ingressou no Ensino Superior deixa de realizar sua matrícula sem comunicar à instituição os motivos do seu afastamento ou cancela definitivamente a mesma. A transferência de instituição ou mudança de curso não se enquadra nessa definição. A definição de Baggi e Lopes [4] é mais ampla e trata evasão como a saída do aluno da IES antes da conclusão de seu curso. Segundo a Secretaria de Educação Superior (SESu ⁵) por meio da Comissão Especial de Estudos sobre a Evasão nas Universidades Brasileiras, a evasão pode ser vista como o desligamento do curso superior (abandono), a transferência, o trancamento ou exclusão pela IES e saída definitiva ou temporária do sistema. Guidotti e Verdum [18] entendem como evasão os movimentos de desistência voluntária dos alunos matriculados, durante qualquer momento do curso, que sejam registrados por abandono, cancelamento, trancamento ou transferência de matrícula para outra instituição.

Santos et al. [15] explicam que existe uma complexidade acerca do tema evasão pela falta de consenso na literatura de uma definição única para evasão, tornando necessário que toda pesquisa feita delimite o escopo do objeto de estudo. Para Lobo [21], a primeira coisa a ser feita ao estudar a Evasão do Ensino Superior é ter clareza e explicitar qual tipo de evasão será tratada: evasão do curso, evasão da IES ou evasão do sistema. Silva Filho et al. [32] consideram que a evasão também pode ser entendida sob dois aspectos: evasão anual média (mede a porcentagem de alunos que, estando matriculados em um sistema de ensino, em uma IES ou em um curso e não se formou, não realizou a matrícula no período seguinte.) e evasão total (mede a quantidade de alunos que, tendo entrado em um determinado curso, IES ou sistema de ensino, não obteve o diploma ao final de um certo número de anos). Já para Lima e Costa [20], a evasão pode ser dividida em relação à duração da evasão, se é definitiva ou temporária, e em relação à amplitude, se é do sistema de ensino superior, da instituição ou do curso.

Para Silva Filho et al. [32], a evasão, além de afetar os resultados do sistema educacional, pode ser considerada desperdício social, acadêmico e econômico. Esse problema afeta tanto instituições da rede privada quanto das públicas, ocasionando uma ociosidade de professores, funcionários, equipamentos e espaço físico, sem contar os gastos de investimento público sem retorno, no caso de IES públicas, e perda de receita para as instituições privadas. Por conta disso, os motivos da evasão estão sendo estudados de forma considerada em diversos países, pois esse problema não se restringe apenas ao Brasil. Na seção seguinte serão detalhados alguns estudos nacionais e internacionais dessa área.

2.3 Estudos sobre evasão

Como citado na seção anterior, o tema da evasão no ensino superior tem recebido um certo destaque em pesquisas por todo o mundo. Os trabalhos feitos por Morosini et

⁵portal.mec.gov.br/sesu

al. [23] e por Santos [14] apresentam um levantamento de artigos já publicados na área no Brasil com intuito de fazer uma análise geral sobre as informações deles. Esses dois estudos mostram a diversidade de possibilidades de pesquisas dentro dessa área.

O trabalho de Morosini et al. [23] apresenta uma revisão bibliográfica dos textos publicados nas principais revistas brasileiras de Educação com classificação A e B, segundo os critérios de qualidade do sistema Qualis ⁶ da Coordenação de Aperfeiçoamento de Pessoal do Nível Superior (Capes ⁷). Durante a análise, foi verificado que a maioria dos estudos de outros autores foi feita de forma qualitativa com base em entrevistas, análise documental e análise bibliográfica. Os autores encontraram oito principais motivos para evasão no ensino superior, dentre eles, aspectos financeiros, interpessoais, baixo desempenho em atividades econômicas, aspectos sociais e baixo nível de motivação. Outra conclusão do trabalho foi que, apesar de todos os estudos analisados falarem que a evasão acarreta uma série de problemas, esses problemas não foram elucidados nos textos.

De forma semelhante, o trabalho de Santos [14] faz uma análise exploratória qualitativa das pesquisas publicadas no Portal de Teses da Capes e nas reuniões semanais da Associação Nacional de Pós-Graduação e Pesquisa em Educação (ANPEd ⁸). Como principais motivos para a evasão no ensino superior, eles citaram: falta de motivação, falta de hábito e técnicas de estudo individualizado, dificuldade de organizar o tempo disponível, conciliar estudo e trabalho e formação escolar anterior precária.

Além de trabalhos relacionados com revisão bibliográfica, estão sendo aplicadas técnicas para estudar casos específicos de evasão em determinadas universidades. Como exemplo, Casartelli et al. [7] estudaram os motivos da evasão na Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS). O estudo foi de caráter qualitativo exploratório em que foi criado um grupo focal com estudantes que evadiram da PUCRS para, com base em um roteiro semi-estruturado, compreender as principais causas da evasão da Universidade. Esse método auxiliou a entender os motivos da evasão de forma mais efetiva do que com o método tradicional utilizado na Universidade (preenchimento de formulário sobre o motivo da evasão), pois os alunos evitam informar o real motivo de sua saída, alegando apenas ser um motivo pessoal.

Com o mesmo objetivo, Perdomo et al. [25] realizaram um estudo dos motivos de evasão na Escola Superior de Ciências da Saúde (ESA) da Universidade do Estado do Amazonas (UEA). Diferentemente do trabalho de Casartelli et al. [7], as informações foram obtidas a partir de documentos legais da própria Universidade, como documentos referentes ao desligamento de alunos matriculados em um período e lista com o total de vagas preenchidas independente da forma de ingresso na Universidade. Os dados foram organizados em uma planilha e receberam um tratamento de estatística descritiva, fornecendo gráficos e tabelas.

As conclusões em todos os trabalhos apresentados ainda são muito inconclusivas. É consenso entre todos os autores citados que analisar dados para entender os motivos da

⁶Qualis é o conjunto de procedimentos utilizados pela Capes para estratificação da qualidade da produção intelectual dos programas de pós-graduação. A classificação de periódicos é realizada pelas áreas de avaliação e passa por processo anual de atualização. Esses veículos são enquadrados em estratos indicativos da qualidade - A1 (o mais elevado) até C (com peso zero). Fonte: Capes

⁷<http://capes.gov.br>

⁸<http://www.anped.org.br>

evasão ainda é uma atividade muito complexa mas que está com avanço nos resultados devido ao aumento de publicações na área nos últimos anos.

Como resultado da análise das pesquisas realizadas, é possível perceber que os dados são obtidos, em geral, de duas formas: por meio de questionários ou por meio de dados que a instituição possui.

Um estudo sobre a forma de obter esses dados de evasão de uma instituição foi realizado por Sarker et al. [30]. Os autores apresentam a situação da evasão universitária no Reino Unido que, assim como no Brasil, está se tornando um problema cada vez maior e gerando diversos custos. Atualmente, o método de entender os motivos da evasão é por meio de questionários. Eles apresentam algumas dificuldades no uso desse método, relacionadas com o custo e esforço de administrar a aplicação e resultados e baixa taxa de participação dos estudantes no preenchimento dos questionários. Os autores propuseram alterar essa forma de avaliação: explorar as informações dos bancos de dados institucionais e de dados abertos para prever os fatores relacionados com evasão e retenção de alunos nas instituições. Foi apresentada uma arquitetura para obtenção, manipulação e exibição dos dados que apresentou resultados bem satisfatórios.

Tendo como base a pesquisa de Sarker et al. [30], surge o questionamento sobre a possibilidade de utilizar dados que a instituição já possui para estudar, de forma mais simples e econômica, o problema da evasão.

2.4 Bacharelado em Ciência da Computação - UnB

O BCC – UnB [11] forma profissionais necessários para diversas atividades humanas, demanda que aumenta cada vez mais devido ao crescimento tecnológico.

O estudante do curso precisa ter afinidade com as Ciências Exatas e criatividade para pensar e implementar soluções. O curso apresenta uma base geral em Matemática, Física, Inglês e Português e uma introdução à áreas específicas da computação no primeiro ano. Nos anos seguintes da graduação, ocorre um reforço na parte de sistemas de computação. Para auxiliar no andamento do curso, o Departamento de Ciência da Computação (CIC) conta com sete laboratórios para a prática profissional e também oferece possibilidade de participação em projetos de pesquisa.

Ao se formar, o estudante poderá atuar nas áreas de desenvolvimento de software, gerência de redes e manutenção/suporte.

2.4.1 Informações sobre o curso

Nome do curso: Ciência da Computação
Habilitação: Bacharelado em Ciência da Computação
Unidade Acadêmica: Instituto de Ciências Exatas (IE)
Campus: Darcy Ribeiro
Turno: Diurno
Reconhecido pelo MEC: sim
Currículo vigente em: 2007/2
Máximo de créditos por período: 32
Mínimo de créditos por período: 15
Créditos exigidos: 240

Limite mínimo de permanência semestral: 7
Limite máximo de permanência semestral: 14
Recomendação de permanência semestral: 9

2.4.2 Disciplinas obrigatórias

A Tabela 2.5 apresenta a lista das disciplinas obrigatórias do BCC – UnB e a quantidade de créditos de cada disciplina.

Tabela 2.5: Disciplinas obrigatórias do BCC – UnB [10].

Código	Disciplina	Créditos
113107	ÁLGEBRA 1	004
113123	ÁLGEBRA LINEAR	006
116378	BANCOS DE DADOS	004
113034	CÁLCULO 1	006
113042	CÁLCULO 2	006
113051	CÁLCULO 3	006
113417	CÁLCULO NUMERICO	004
116351	CIRCUITOS DIGITAIS	006
116301	COMPUTAÇÃO BÁSICA	006
116441	ENGENHARIA DE SOFTWARE	004
116319	ESTRUTURAS DE DADOS	004
118001	FÍSICA 1	004
118010	FÍSICA 1 EXPERIMENTAL	002
118028	FÍSICA 2	004
118036	FÍSICA 2 EXPERIMENTAL	004
118044	FÍSICA 3	004
118052	FÍSICA 3 EXPERIMENTAL	004
145971	INGLÊS INSTRUMENTAL 1	004
116416	INTRO SISTEMAS INFORMAÇÃO	004
113930	INTRODUÇÃO A TEORIA DOS GRAFOS	004
140481	LEITURA E PRODUÇÃO DE TEXTOS	004
116343	LINGUAGENS DE PROGRAMAÇÃO	004
116394	ORG ARQ DE COMPUTADORES	004
116327	ORGANIZAÇÃO DE ARQUIVOS	004
115045	PROBABILIDADE E ESTATÍSTICA	006
113956	PROGRAMAÇÃO SISTEMÁTICA	004
116467	SISTEMAS OPERACIONAIS	006
116432	SOFTWARE BASICO	004
113115	TEORIA DOS NÚMEROS	004
116459	TRADUTORES	006

Além dessas matérias, os estudantes precisam concluir as Cadeias e Ciclos apresentados nas Tabelas 2.6, 2.7, 2.8, 2.9 e 2.10.

Tabela 2.6: Cadeia 4: Ciclo 3: o aluno deverá cursar [10].

Código	Disciplina		Créditos
113948	LINGUAGENS FORMAIS E AUTÔMATOS	E	004
116360	TEORIA DA COMPUTAÇÃO	OU	004
116882	AUTÔMATOS E COMPUTABILIDADE		006

Tabela 2.7: Cadeia 5: Ciclo 3: o aluno deverá cursar [10].

Código	Disciplina		Créditos
116475	TRABALHO DE GRADUAÇÃO	OU	006
116912	TRABALHO DE GRADUAÇÃO 1	E	002
116921	TRABALHO DE GRADUAÇÃO 2		004

Tabela 2.8: Cadeia 6: Ciclo 3: o aluno deverá cursar [10].

Código	Disciplina		Créditos
137481	LÓGICA 1	OU	006
117366	LÓGICA COMPUTACIONAL 1		004

Tabela 2.9: Cadeia 7: Ciclo 3: o aluno deverá cursar [10].

Código	Disciplina		Créditos
113859	ANÁLISE DE ALGORITMOS	OU	004
117536	PROJETO ANÁLISE DE ALGORITMOS		004

Tabela 2.10: Cadeia 8: Ciclo 3: o aluno deverá cursar [10].

Código	Disciplina		Créditos
116424	TRANSMISSÃO DE DADOS	OU	004
204315	TELEINFORMÁTICA E REDES 1		004

As ementas das disciplinas obrigatórias e seus pré-requisitos estão mostradas no Anexo I.

2.4.3 Disciplinas optativas

A Tabela 2.11 apresenta a lista das disciplinas optativas do Bacharelado em Ciência da Computação e a quantidade de créditos de cada disciplina.

Tabela 2.11: Disciplinas optativas do Bacharelado em Ciência da Computação da UnB [10].

Código	Disciplina	Créditos
113131	ALGEBRA 2	004
117145	ALGEBRA 3	004
113204	ANALISE 1	006
113212	ANALISE 2	004
117137	ANALISE 3	004
113972	ANALISE COMBINATORIA	004
145491	ANALISE DA IMAGEM	004
117188	ANALISE E PROJETO DE SISTEMAS	006
153206	ANALISE GRAFICA 1	004
113506	ANALISE NUMERICA 1	004
194671	APREND E DESENV DO PNEE	004
116594	ARQUITETURAS AVANÇADAS	004
316407	ARQUITETURAS VLSI	004
116769	ARTES GRAF E TIPOGRAF DIGITAL	004
192287	AVALIACAO DA APRENDIZAGEM	004
204358	AVALIAÇÃO DESEMPENHO DE REDES	004
113824	CALCULO DE PROBABILIDADE 1	006
113832	CALCULO DE PROBABILIDADE 2	004
204340	COMPUT MÓVEL REDES SEM FIO	004
186163	COMUNICACAO E NEGOCIACAO	004
117285	CONSTRUCAO DE PECAS MULTIMIDIA	004
181919	CONTAB E ANALISES DE BALANCOS	004
181129	CONTABILIDADE GERAL 1	004
192104	CURRICULOS E PROGRAMAS 1	004
117561	DEPEND EM SIS COMPUTACIONAIS	004
153133	DESENHO GEOMETRICO	006
116777	EDSITORACAO POR COMPUTADOR	004
191621	EDUC AMB E PRAT COMUNITARIAS	004
192562	EDUCACAO A DISTANCIA	004
153001	ELEM LING ESTÉTICA HIST ARTE 1	004
153010	ELEM LING ESTÉTICA HIST ARTE 2	004
200662	EMPREENDEDORISMO I TECNOLÓGICA	004
170721	EMPRESA JUNIOR 1	006
170739	EMPRESA JUNIOR 2	006
113301	EQUACOES DIFERENCIAIS 1	004
124664	ERGONOMIA 1	006
116661	ESTU INTELIGENCIA ARTIFICIAL	004
116556	ESTUDOS EM CRIPTOGRAFIA	004
116629	ESTUDOS EM SIS DE COMPUTACAO	004
116521	ESTUDOS PROCES ESPECTRAL	004

Continua na página seguinte

Tabela 2.11 – *Continuação da tabela da página anterior*

Código	Disciplina	Créditos
116734	ESTUDOS SISTEMAS INFORMACAO	004
153079	EXPRES SUPERF VOL, MOVIMENTO	004
116688	FERRAMENTA AN SISTEMAS	004
137634	FILOSOFIA DA LINGUAGEM	004
139602	FILOSOFIA DA MENTE	004
118061	FISICA 4	004
118079	FISICA 4 EXPERIMENTAL	004
140511	FONETICA FONOLOGIA PORTUGUES	004
200379	FUND COMPUT DE ROBÓTICA	004
153699	FUND DA LINGUAGEM VISUAL	006
117579	FUND DE SISTEMAS INTELIGENTES	004
316377	FUND SIS COMPUTACIONAIS	004
191663	FUNDAMENTOS DA EDUC AMBIENTAL	004
117161	GEOMETRIA 1	004
117170	GEOMETRIA 2	004
113328	GEOMETRIA DIFERENCIAL 1	004
204331	GER E SEGURANÇAO DE REDES	004
116700	GERENCIA DE PROJETOS	004
117200	GERENCIA DE REDES	004
186171	GESTAO ORGANIZACIONAL	004
152901	INFOESTETICA 1 - EST IMAG INF	004
116858	INFORMATICA APLIC A EDUCACAO	004
116726	INFORMATICA E SOCIEDADE	002
116718	INFORMATICA NAS ORGANIZACOES	004
142573	INGLÊS INSTRUMENTAL 2	004
316024	INTELIGÊNCIA ARTIFICIAL 1	004
316032	INTELIGÊNCIA ARTIFICIAL 2	004
117587	INTERAÇÃO HUM COMPUTADOR	004
170054	INTR A ATIVIDADE EMPRESARIAL	004
117129	INTRO A COMPUTACAO ALGEBRICA	004
113433	INTRO A PROGRAMACAO LINEAR	004
117072	INTRO AS EQUACOES DIF PARCI	004
116653	INTRO INTELIGENCIA ARTIFICIAL	004
117544	INTROD A SISTEMAS MULTIAGENTES	004
117552	INTROD AO DESENV DE JOGOS	004
181013	INTRODUCAO A ADMINISTRACAO	004
185035	INTRODUÇÃO A CIÊNCIA POLÍTICA	004
132012	INTRODUÇÃO A ECONOMIA	004
191019	INTRODUCAO A EDUCACAO	004
191299	INTRODUCAO A EDUCACAO ESPECIAL	004
153061	INTRODUCAO A ESCULTURA	006
137553	INTRODUCAO A FILOSOFIA	004
140082	INTRODUÇÃO A LINGÜÍSTICA	004

Continua na página seguinte

Tabela 2.11 – *Continuação da tabela da página anterior*

Código	Disciplina	Créditos
156272	INTRODUCAO A PINTURA	006
124010	INTRODUÇÃO A PSICOLOGIA	004
134465	INTRODUÇÃO A SOCIOLOGIA	004
117242	INTRODUCAO AO PROC DE IMAGENS	004
156264	INTRODUCAO AOS MULTI MEIOS	006
116491	INTRODUCAO COMPUTACAO SONICA	004
153711	INTRODUCAO PROGRAMACAO VISUAL	006
116670	LEVANTAMENTO DADOS PESQUISA	004
116815	LING P/INTELIGENCIA ARTIFICIAL	004
147630	LINGUA CHINESA 1	004
147648	LÍNGUA CHINESA 2	004
147656	LINGUA CHINESA 3	004
150649	LÍNGUA SINAIS BRAS - BÁSICO	004
316041	LINGUAGENS DECLARATIVAS	004
116751	LINGUAGENS P/ SIS EDITORACAO	004
117374	LÓGICA COMPUTACIONAL 2	004
117358	LÓGICA MAT E COMPUTACIONAL	004

Capítulo 3

Mineração de Dados

Nesse capítulo, são descritos os conceitos gerais de mineração de dados necessários ao entendimento desse trabalho. Na Seção 3.1 é apresentada uma visão geral sobre a área de mineração de dados e, na Seção 3.2, os principais conceitos da área. A Seção 3.3 descreve os diferentes tipos de aprendizado. As técnicas de mineração são apresentadas na Seção 3.4. A Seção 3.5 apresenta ferramentas utilizadas para mineração de dados.

3.1 Apresentação

Segundo Han e Kamber [19], a mineração de dados começou a ter sua importância percebida nos últimos anos pela indústria da informação e pela sociedade como um todo devido à grande quantidade de dados existentes e à necessidade de transformar esses dados em informações úteis e em conhecimento.

De acordo com Witten et al. [35], essa grande quantidade de dados existentes aumenta a cada dia e não parece ter um fim. Alguns motivos citados que originam nessa quantidade de dados são: facilidade de salvar dados nos computadores, baixo custo de discos rígidos e as opções de armazenamento *online*, elementos da computação ubíqua que armazenam diversos tipos de dados, etc.

É possível perceber, segundo Witten et al. [35], o crescimento de uma lacuna entre a geração de dados e o entendimento deles. A geração e o entendimento de dados são inversamente proporcionais, ou seja, à medida que a quantidade de dados existente aumenta, o entendimento desses dados decresce.

Escondidas nesse grande volume de dados, existem informações úteis que raramente são explícitas ou utilizadas. A mineração de dados se preocupa em identificar padrões nesses dados armazenados eletronicamente de forma automatizada por um computador.

3.2 Conceitos básicos

Essa seção apresenta os principais conceitos que precisam ser compreendidos para estudo de mineração de dados.

3.2.1 Mineração de dados e Extração de conhecimento

Han e Kamber [19] definem, de forma ampla e simples, mineração de dados como o processo de extrair ou, como o próprio nome diz, “minerar” conhecimento de grandes quantidades de dados.

No entanto, segundo Fayad et al. [16], a mineração está além disso. A mineração de dados é uma das etapas do *Knowledge-Discovery in Databases* (KDD ou extração de conhecimento, em português) que consiste na aplicação de algoritmos específicos para extrair padrões dos dados.

Por conta disso, o conceito de extração de conhecimento costuma ser usado como sinônimo de mineração de dados, o que é um erro pois representam coisas distintas. A extração de conhecimento é, como pode ser visto na Figura 3.1, um processo iterativo de identificação de informações válidas, úteis e legíveis nos dados, mesmo que somente após a realização de alguns processamentos neles, e também de determinação de novos padrões.

Como mostrado na Figura 3.1, o processo de extração de conhecimento é composto por sete etapas:

1. Limpeza dos dados

Remoção de dados inconsistentes ou sem significado.

2. Integração dos dados

Combinação de dados de diferentes bancos de dados.

3. Seleção de dados

Recuperação de dados dos bancos de dados que são relevantes para a aplicação.

4. Transformação dos dados

Aplicação de operação de resumo ou agregação com intuito de transformar ou consolidar os dados em formas adequadas para a mineração.

5. Mineração de dados

Processo essencial de aplicação de métodos inteligentes para extração de padrões de dados.

6. Avaliação dos padrões

Identificação dos padrões que realmente são interessantes para representação do conhecimento baseado em mensurações.

7. Apresentação do conhecimento

Uso de técnicas de visualização de dados e representação do conhecimento para apresentação para o usuário dos conhecimentos que foram minerados.

Para Han e Kamber [19], a mineração de dados envolve técnicas multidisciplinares, como tecnologias de banco de dados e de *data warehouse*, estatística, *machine learning*, computação de alta performance, reconhecimento de padrões, redes neurais, visualização de dados, recuperação de informações, processamento de imagens e sinais e analisadores espaciais e temporais de dados.

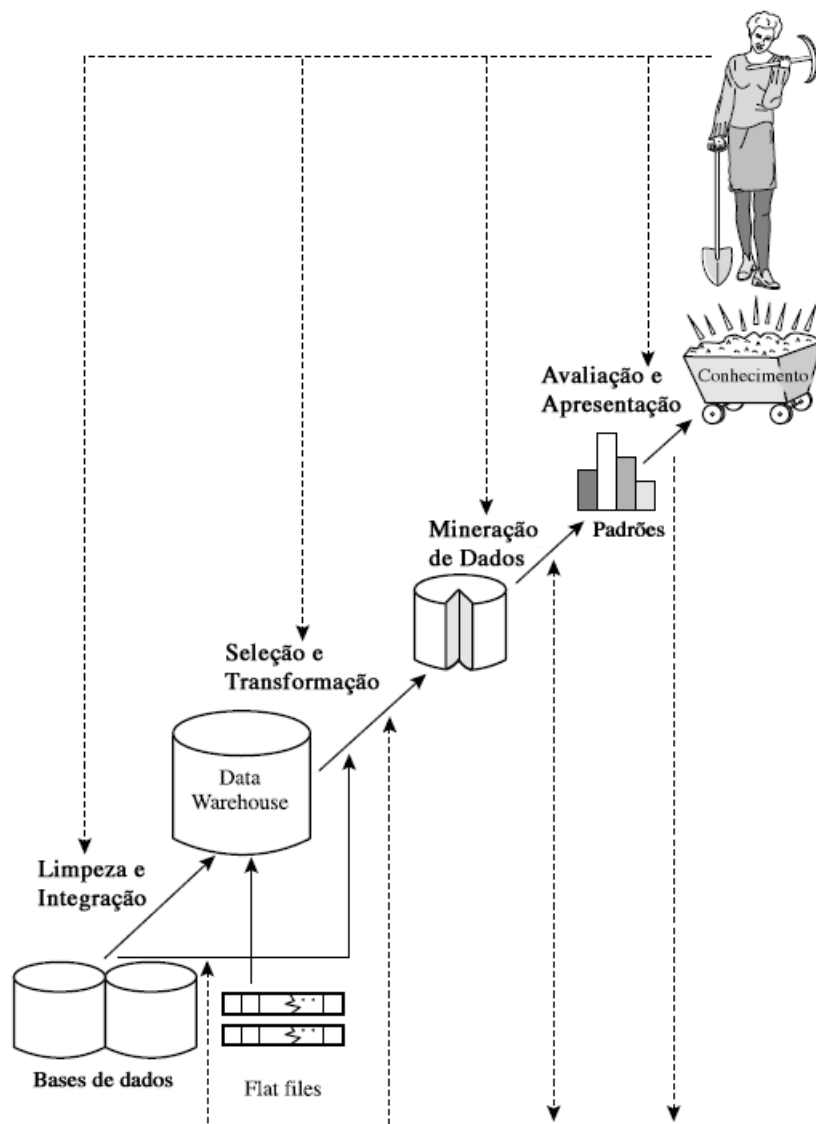


Figura 3.1: Etapas do processo de extração de conhecimento. Adaptado de Han e Kamber [19].

É importante ressaltar que a mineração de dados deve ter como base técnicas eficientes e escaláveis. Um algoritmo é escalável quando o tempo de execução aumenta de forma linear proporcionalmente ao tamanho dos dados de acordo com os recursos disponíveis, como memória e espaço em disco.

Os conhecimentos interessantes, regularidades ou informações de alto nível extraídas dos bancos de dados com mineração de dados podem ser vistas de diferentes ângulos. O conhecimento obtido pode ser usado em diferentes situações, como: tomada de decisão, controle de processo, gerenciamento de informações e processamento de consultas.

Por conta disso, a mineração de dados é considerada uma das áreas mais importantes em sistemas de banco de dados e de informações, é uma das áreas de desenvolvimento interdisciplinar mais promissoras da tecnologia da informação.

3.2.2 Objetivos da extração de conhecimento

De acordo com Fayad et al. [16], existem dois tipos de objetivos no processo de extração de conhecimento: verificação e descoberta.

- Verificação

O sistema se limita a verificar as hipóteses definidas pelos usuários.

- Descoberta

O sistema automaticamente encontra novos padrões nos dados.

O objetivo de descoberta pode ser dividido em dois outros objetivos: predição e descrição.

- Predição

O sistema encontra padrões para prever o comportamento futuro de algumas entidades.

- Descrição

O sistema encontra padrões para apresentar aos usuários de forma compreensível por humanos.

3.2.3 Arquitetura de um sistema de mineração de dados

A arquitetura típica de um sistema de mineração de dados, de acordo com Han e Kamber [19], é apresentada na Figura 3.2.

A arquitetura básica de um sistema de mineração de dados possui diversos componentes, descritos a seguir.

- Banco de dados, *Data Warehouse*, *World Wide Web* e outros repositórios de informações.

Um ou mais conjuntos de bancos de dados, *Data warehouses* ou outros tipos de repositórios de informações, como planilhas; é nesse componente que as técnicas de limpeza e de integração de dados costumam ser aplicadas.

- Banco de dados ou servidor de *Data warehouse*

Esses componentes são responsáveis por determinar os dados que são relevantes, de acordo com as requisições do usuário do sistema de mineração de dados.

- Base de conhecimento

Domínio que é usado para guiar as pesquisas ou avaliar quais os padrões resultantes que são interessantes para a aplicação.

Esse conhecimento pode incluir hierarquia de conceitos (usada para organizar os atributos e seus valores em diferentes níveis de abstração), crenças de usuários (usadas para avaliar o quão interessante um padrão é baseado em sua imprevisibilidade), restrições adicionais, metadados, entre outros.

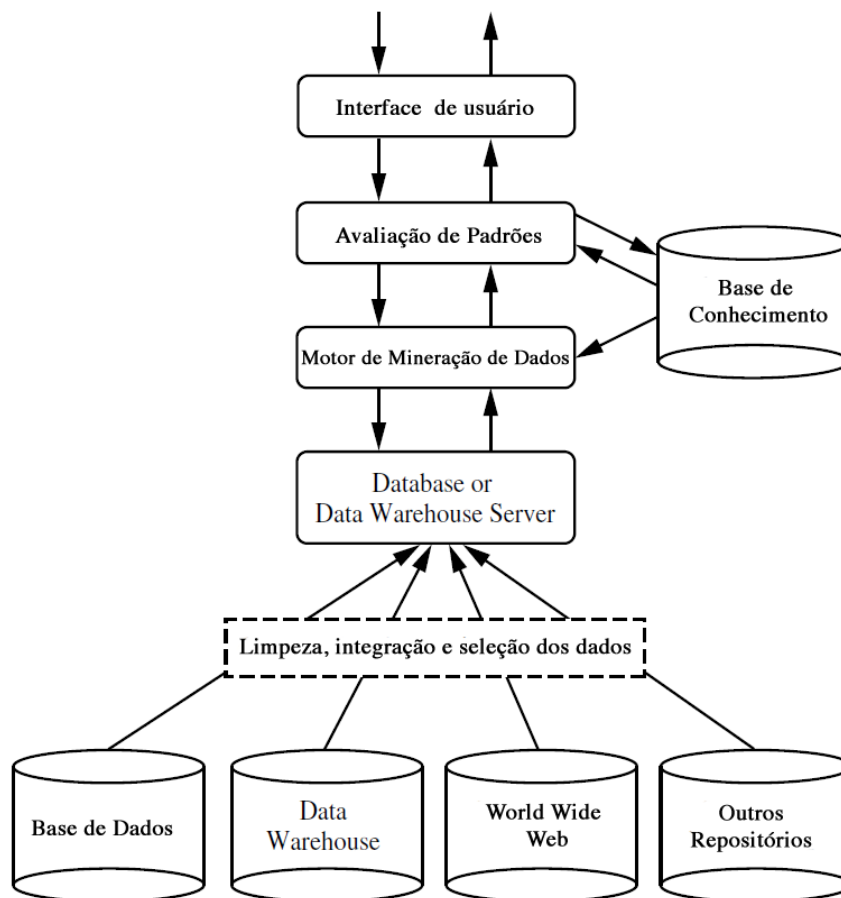


Figura 3.2: Arquitetura típica de um sistema de mineração de dados. Adaptado de Han e Kamber [19].

- Motor de mineração de dados

Esse componente é essencial para o sistema de mineração de dados. Ele consiste em um conjunto de módulos funcionais para realização de atividades como caracterização, associação e correlação estatística, classificação, predição, clusterização, entre outros.

- Módulo de avaliação de padrões

Aplicação de medidas para determinar o grau de interesse de um padrão e é responsável pela interação com os módulos de mineração de dados, de modo a pesquisar padrões interessantes.

Esse módulo pode ser integrado ao de mineração, dependendo do método escolhido para a implementação. Visando a eficiência da mineração de dados, é altamente recomendável inserir o avaliador de interesse dos padrões o mais profundamente possível no processo de mineração, a fim de limitar as buscas apenas aos padrões interessantes.

- Interface do usuário

Módulo responsável por realizar a comunicação entre os usuários e o sistema de mineração de dados, permitindo que o usuário interaja com o sistema especificando consultas ou atividades, fornecendo informações para deixar a busca mais focada, explorando a mineração de dados com base nos resultados intermediários. Além disso, a interface permite que o usuário possa navegar por esquemas de banco de dados e de *data warehouse* ou por estrutura de dados, avaliar os padrões que foram minerados e visualizá-los de diferentes formas.

3.3 Tipos de aprendizado

De acordo com Han e Kamber [19], a mineração de dados envolve técnicas de *machine learning* e, portanto, é importante entender os tipos de aprendizado que existem.

Alpaydin [2] classifica o aprendizado em três tipos: supervisionado, não supervisionado e por reforço, detalhados nas subseções seguintes.

3.3.1 Aprendizado supervisionado

O objetivo do aprendizado supervisionado é aprender a fazer um mapeamento dos valores de entrada para os de saída com base em valores corretos fornecidos por um supervisor.

Para isso, conforme explicado por Donalek [13], o processo de aprender possui uma fase de aprendizado com dados de testes, chamada de treinamento. São fornecidos os dados de entrada e resultados esperados e alguns exemplos de dados já possuem resultado correto determinado pelo usuário. Também é reforçada a importância da generalização, que é a habilidade de produzir saídas razoáveis para entradas que não foram encontradas na fase de treinamento.

3.3.2 Aprendizado não supervisionado

Ao contrário do aprendizado supervisionado, no não supervisionado não existe um supervisor, apenas existem dados de entrada. O objetivo é encontrar regularidades nos dados de entrada.

Pujari [26] define o aprendizado não supervisionado como a descoberta automática (sem intervenção humana) de informações escondidas nos dados por meio de identificação de padrões.

3.3.3 Aprendizado por reforço

O aprendizado por reforço é interessante em aplicações cuja saída do sistema seja uma sequência de ações e, nesse caso, o que importa é a política definida pelo conjunto de ações corretas para atingir um determinado objetivo. Uma única ação não é importante e não existe uma ação que seja melhor do que as outras em um estado intermediário, o que torna uma ação boa é se ela faz parte de uma política que levará ao alcance do objetivo.

No aprendizado por reforço, então, o programa deverá aprender com base nas ações corretas e incorretas realizadas no passado para criar uma boa política.

3.4 Técnicas de mineração de dados

Existem diversas técnicas de mineração de dados: classificação, regressão, clusterização, sumarização, associação, redução, entre outras. Para cada uma dessas técnicas, já existem diversos algoritmos desenvolvidos.

As três técnicas que serão focadas no desenvolvimento desse trabalho são: classificação, regressão e clusterização, explicadas a seguir.

3.4.1 Classificação

A classificação é uma técnica baseada no aprendizado supervisionado. De forma geral, explicada por Donalek [13], essa técnica divide as amostras de dados em classes determinadas pelo usuário e utiliza os dados de treino, que já foram classificados, para classificar o restante dos dados nas classes corretas.

Han e Kamber [19] descrevem a técnica como um processo de duas etapas. A primeira etapa, chamada de fase de treino, consiste na criação de um classificador que descreve um conjunto de classes e conceitos pré-determinados. O algoritmo de classificação irá construir esse classificador por meio do aprendizado baseado no conjunto de testes, representados por uma tupla (vetor de atributos de dimensão n) que conterá atributos e pertencerá a uma classe já determinada por outros atributos do banco de dados, chamados atributos rótulos de classes. A Figura 3.3 representa um exemplo dessa primeira etapa.

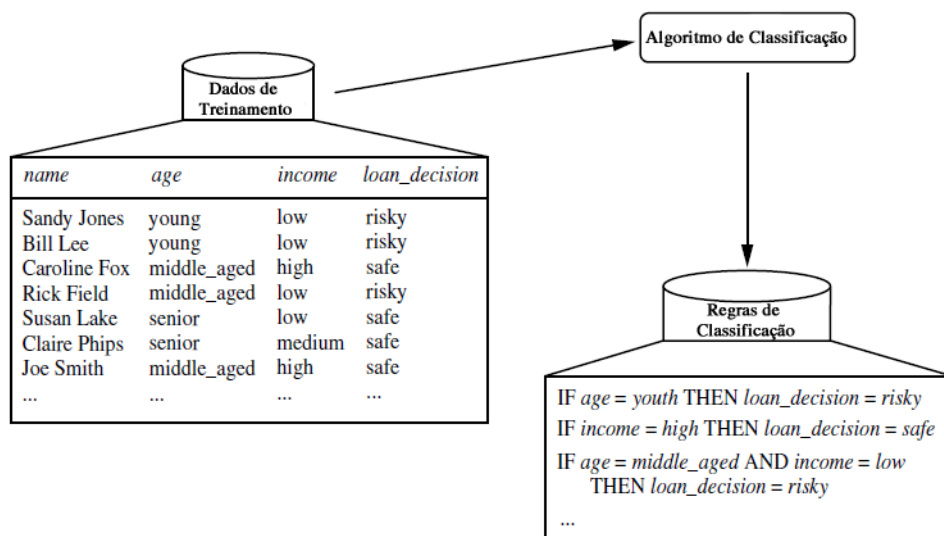


Figura 3.3: Exemplo da etapa de treinamento da classificação, segundo Han e Kamber [19]. Nesse exemplo, a classificação terá objetivo de determinar se um empréstimo é de risco ou é seguro. Os arquivos de entrada já possuem exemplos de algumas pessoas e, com base nesses dados, o algoritmo de classificação determina as regras da classificação.

A segunda etapa é onde ocorre efetivamente a classificação. Um ponto importante dessa técnica é estimar a acurácia da classificação. Como não é possível determinar a acurácia do classificador unicamente com base na fase de treino, uma vez que podem ter ocorrido anomalias, ou seja, dados com padrões diferentes dos demais, nos arquivos de treinos que não correspondem aos demais dados do banco, é criado um conjunto de

testes, formado por tuplas com o atributo rótulo de classe já associado e que não estavam presentes na fase de treino. A acurácia será dada pela porcentagem de tuplas do conjunto de testes que foram classificadas corretamente pelo classificador, ou seja, que o atributo rótulo de classe da tupla seja igual ao da predição feita pelo classificador. Caso essa acurácia seja considerável aceitável pela aplicação, é possível utilizar esse classificador para prever qual a classe das tuplas em que não se conhece o atributo rótulo de classe. A Figura 3.4 representa um exemplo dessa segunda etapa, relacionado ao exemplo dado na Figura 3.3.

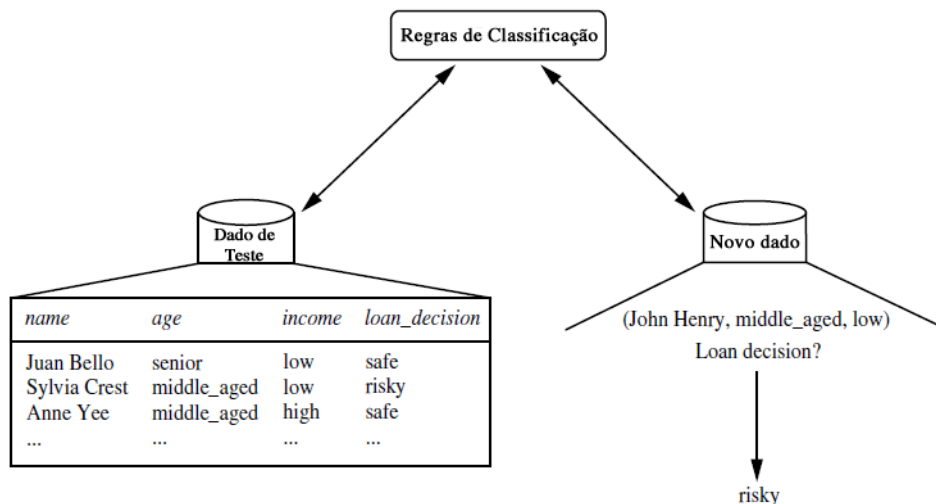


Figura 3.4: Exemplo da segunda etapa da classificação, segundo Han e Kamber [19]. Esse exemplo representa a segunda etapa do que foi representado na Figura 3.3. Os arquivos de testes, que já estão classificados mas não foram utilizados na fase de treinamento, rodam sobre as regras de classificação para verificar a acurácia do classificador. Caso a acurácia esteja satisfatória para a aplicação, nossos dados iram rodar sobre essas regras de classificação para determinar se o empréstimo é de risco ou não.

Algoritmos de classificação

Existem diversos algoritmos de classificação. Wu et al. [36] apresentaram uma lista com os 10 melhores algoritmos de mineração de dados. Dentre os de classificação, destacam-se: *Naive Bayes*, *Support Vector Machine* (SVM), *k-nearest neighbor* (kNN) e C4.5.

O **Naive Bayes** é considerado de fácil construção e não faz uso de parâmetros iterativos complicados, o que permite que ele seja facilmente aplicado a grandes quantidades de dados. Sua interpretação também é relativamente fácil, inclusive para pessoas sem conhecimentos avançados em computação. Além disso, apesar de sua simplicidade, ela costuma ter bons resultados, podendo ser invocado em situações que exijam robustez. A classificação acontece por meio de cálculos de probabilidade para determinar a probabilidade de cada instância pertencer a uma classe.

O **SVM** é um algoritmo considerado que vale a pena ser testado em aplicações de aprendizado de máquina por ser um dos algoritmos com maior robustez e acurácia. Suas vantagens são o bom fundamento teórico, necessidade de poucos exemplos para a fase de treinamento e funciona independente da quantidade de atributos. Seu funcionamento

ocorre por meio de uma função matemática criada na fase de treinamento capaz de distinguir as possíveis classes.

O **kNN** é um algoritmo que detecta um grupo de k objetos que foram utilizados na fase de treinamento e que mais se assemelha a um novo dado que está sendo testado e classifica essa nova instância de acordo com a classificação que mais ocorre nessa vizinhança. Os três elementos chave desse algoritmo são: o conjunto de registros que foram utilizados na fase de treinamento, a métrica utilizada para computar a distância dos objetos (o que determina a vizinhança) e o valor de k (quantidade de vizinhos). A classificação de uma nova instância ocorre com o cálculo da distância dos objetos já classificados em relação a esse, a identificação dos k vizinhos mais próximos e a determinação da classe por meio da classificação desses vizinhos.

O **C4.5** é um algoritmo baseado em árvore de decisão descendente do ID3, outro algoritmo bem conhecido que gera árvore de decisão para classificar as instâncias. O C4.5 constrói a árvore por meio de uma abordagem dividir para conquistar. **ID3**, segundo Adhatrio et al. [1], é um algoritmo matemático inventado em 1979 utilizado para criação de árvores de decisão. O *splitting criteria*, ou seja, o critério utilizado para escolha do atributo em cada nó da árvore, é a entropia. A entropia é a medida de quão aleatório é um conjunto de dados e, portanto, quanto maior a entropia, maior a necessidade de informações para descrever o dado.

3.4.2 Regressão

Assim como a classificação, a regressão também é uma técnica baseada no aprendizado supervisionado. Roiger e Geatz [27] definem a regressão como uma generalização de um conjunto de dados numéricos por meio da criação de equações matemáticas relacionando um ou mais atributos em um único atributo numérico de saída. Apesar de uma regressão poder ser não linear, em geral o uso popular da regressão é para modelo lineares.

Segundo Han e Kamber [19], ao contrário da classificação, que possui, em geral, um resultado booleano como saída, ou seja, se uma tupla pertence ou não a uma classe, a regressão tem como saída um valor numérico.

Alpaydin [2] explica que a técnica é supervisionada já que não se sabe inicialmente qual a fórmula utilizada para fazer a regressão, ela será obtida com base no conjunto de testes. Por meio de equações estatísticas, é possível determinar qual a função que representa a regressão. É importante ressaltar que, apesar de na fase de treino o erro diminuir à medida que o grau da equação aumenta, esse alto grau pode ser interessante apenas para detectar os exemplos fornecidos no treino, mas não a forma geral do restante dos dados. Portanto, é importante tomar cuidado ao determinar o grau da equação.

3.4.3 Clusterização

Ao contrário das técnicas citadas acima, a clusterização é baseada no aprendizado não supervisionado.

De forma geral, definida por Donalek [13], a clusterização agrupa os dados em sub-grupos (*clusters*) de forma que os dados em cada grupo compartilhem características em comum.

3.5 Ferramentas de mineração de dados

Existem diversas ferramentas que facilitam a mineração de dados, como Weka, Rapid Miner ¹ e Matlab ². O Weka, que foi utilizado nesse trabalho, será apresentado brevemente.

Weka, que representa uma abreviação de *Waikato Environment for Knowledge Analysis*, é, segundo o próprio site do projeto [24], uma coleção de algoritmos do estado da arte de *Learning Machine* para a realização de atividades de mineração de dados.

De acordo com Witten et al. [35], o Weka foi desenvolvido na Universidade de Waikato, Nova Zelândia, utilizando a linguagem Java e compatível com Linux, Windows e Macintosh. Ele provê uma interface para diferentes algoritmos de aprendizado de máquina, métodos para pré e pós-processamento de dados e para avaliação dos resultados dos esquemas de aprendizado em qualquer conjunto de dados fornecido como entrada.

O Weka é uma ferramenta que facilita o processo de mineração de dados. Dentre as vantagens, podemos destacar a existência dos principais algoritmos de mineração, ou seja, classificação, clusterização, associação e seleção de atributos. Além disso, ele foi desenvolvido de forma que possibilite o uso rápido dos métodos existentes de forma flexível em novos conjuntos de dados. Ele fornece suporte para todo o processo de mineração de dados, que vai desde a preparação dos dados de entrada até a visualização do resultado do aprendizado. Também estão inclusos uma variedade de ferramentas de pré-processamento. Existe uma facilidade em comparar os métodos existentes e identificar o mais apropriado para cada problema. Todas essas etapas de fazer um pré-processamento em um conjunto de dados, organizar os dados em esquemas de aprendizado e analisar o resultado e a performance de cada classificador podem ser realizadas sem escrever nenhuma linha de código.

Existem três formas de utilizar o Weka. A primeira é aplicar um dos métodos de aprendizado a um conjunto de dados e analisar o resultado da saída. A segunda é utilizar os modelos de aprendizado para gerar predições em novas instâncias. E, por último, a terceira é aplicar diferentes formas de aprendizado e comparar o desempenho de cada uma para escolher a melhor a ser usada na predição. Esse desempenho é medido por um módulo de avaliação comum que avalia a performance de todos os classificadores.

O recurso mais valioso que o Weka fornece é a implementação dos esquemas de aprendizado, seguido pelas ferramentas de pré-processamento de dados, os filtros.

Existem três principais interfaces gráficas para usuários no Weka: *Explorer*, *Knowledge Flow* e *Experimenter*. Também é possível utilizar o Weka por linha de comando.

A interface chamada *Explorer*, exibida na Figura 3.5, é o jeito mais fácil de utilizar o Weka e é por ela que os usuários costumam iniciar a utilização da ferramenta. Ela permite acesso a todos os recursos utilizando um menu e formulários, tornando possível a criação de uma árvore de decisão, ou qualquer outro algoritmo existente na ferramenta, com base em um conjunto de dados fornecido como entrada. Essa interface facilita a exploração dos algoritmos existente ao apresentar as opções como escolhas no menu e campos de formulários para preencher. Além disso, ao passar o *mouse* nas opções, uma explicação é dada sobre a funcionalidade. Uma desvantagem dessa interface é que ela mantém todos

¹<https://rapidminer.com/>

²<http://www.mathworks.com/products/matlab/>

os dados em memória e, por isso, a *Explorer* só pode ser aplicada a programas de pequeno e médio porte.

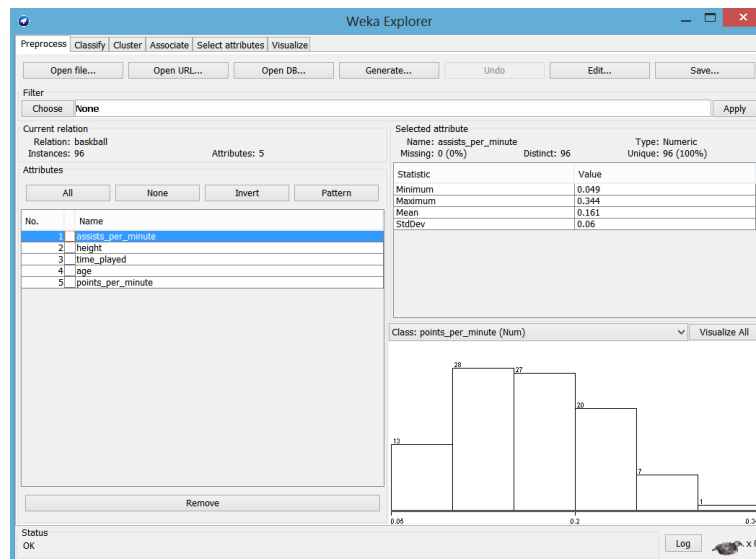


Figura 3.5: Interface *Explorer* do Weka versão 3.6 no Windows.

Knowledge Flow é uma interface que permite o projeto de configurações para o processamento de dados. Essa interface, representada na Figura 3.6, permite análises mais complexas por meio da combinação de componentes que representam fontes de dados, ferramentas de pré-processamento, algoritmos, métodos de avaliação e módulos de visualização. Essa combinação é facilitada pelo recurso de arrastar caixas que representam os algoritmos e fontes de dados e juntá-las como desejar.

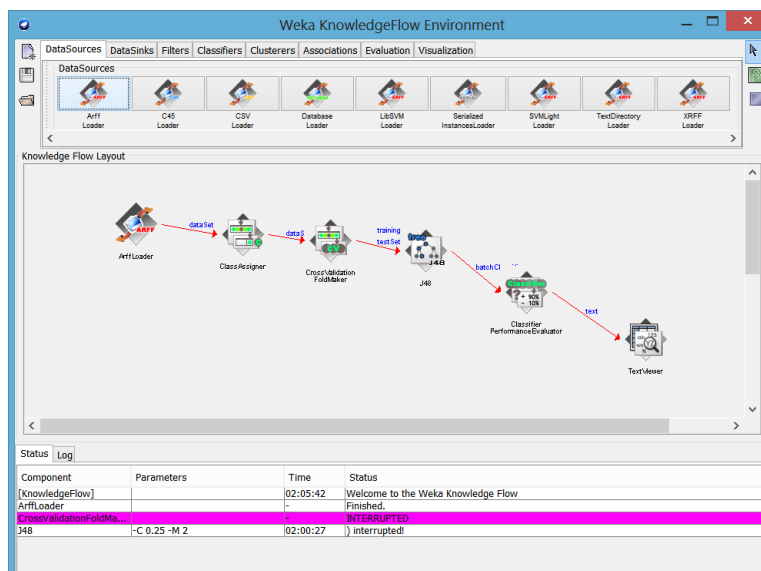


Figura 3.6: Interface *Knowledge Flow* do Weka versão 3.6 no Windows.

A última interface gráfica, representada na Figura 3.7, é chamada de *Experimenter* e tem o objetivo de responder questões básicas quando são aplicadas técnicas de classifica-

ção e regressão. Uma pergunta que essa interface ajuda a responder é: quais os métodos e valores de parâmetros que possuem melhor resultado para um determinado problema? Essa ajuda é feita pela facilidade que a interface provê de executar classificadores e filtros com diferentes configurações de parâmetros de forma automatizada e interativa, armazenando as estatísticas de desempenho e realizando testes significativos.

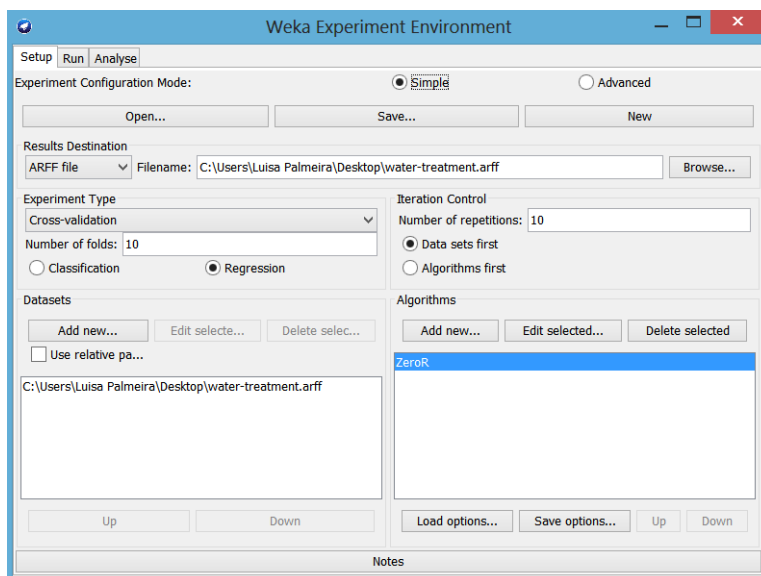


Figura 3.7: Interface *Experimenter* do Weka versão 3.6 no Windows.

Capítulo 4

Trabalhos Relacionados

Esse capítulo apresenta trabalhos sobre evasão que fizeram uso de análise e/ou mineração de dados, com o objetivo de verificar os procedimentos utilizados para esse estudo em instituições com base nessas técnicas, visando auxiliar o desenvolvimento da metodologia utilizada nesse trabalho. A Seção 4.1 apresenta uma visão geral dos trabalhos estudados. A Seção 4.2 detalha os trabalhos realizados na UnB e a Seção 4.3 os realizados no Brasil. A Seção 4.4 compara os dados e metodologias desses trabalhos.

4.1 Visão geral

Como dito nos capítulos anteriores, os trabalhos sobre evasão têm recebido um maior destaque nos últimos anos, avançando nos resultados da análise a cada pesquisa. Esse capítulo explora, de forma mais aprofundada, as metodologias utilizadas nas pesquisas para fazer essas análises, abordando como foram tratados os dados, quais foram os métodos utilizados para extração, manipulação e análise.

De acordo com Araque et al. [3], existem duas maneiras de realizar uma pesquisa sobre a evasão dos estudantes. A primeira delas, mais comum, é a pesquisa com aplicação de questionários e entrevistas. Esse método é utilizado para perceber as questões subjetivas e qualitativas do real motivo que o aluno decidiu por não terminar seus estudos. No entanto, essa metodologia possui algumas dificuldade de implementação, que devem ser trabalhadas para que não inviesse a coleta de dados. Esse tipo de modalidade, por visar obter respostas mais subjetivas, pode se torna muito invasiva na vida do aluno e caso não seja tratada da melhor maneira, pode gerar alguns resultados enviesados. A outra maneira é a análise dos dados dos estudantes que as instituições guardam por meio do ato de matrícula, produzidos pelos estudantes na instituição e armazenados nos sistemas acadêmicos. Por tratar o problema de maneira mais objetiva, essa metodologia se torna menos invasiva e mais transparente, mas acaba não levando em consideração quais os motivos reais da saída do aluno.

Algumas metodologias estudadas apresentam as duas perspectivas, utilizando os resultados da análise previa dos dados para suportar as informações subjetivas coletados nas entrevistas ou como um direcionamento de quais aspectos abarcar na fase de entrevista. O uso dos dois métodos de maneira intercalada, vemos como complementar, com um método dando suporte e reforçando o resultado do outro. No entanto, a maioria dos

estudos utilizaram apenas de análise dos dados dos estudantes, seja com ou sem uso de técnicas de mineração de dados, obtendo resultados satisfatórios.

Em todas as metodologias estudadas, o primeiro passo é definir quais foram os casos de evasão estudados. Segundo Souza et al. [33], as formas de evasão estudadas foram: por tempo médio de curso, por abandono, por desistência oficial, por eliminação, por jubilação, por transferência de instituição ou interna.

4.2 Evasão na UnB

Na UnB já existem estudos sobre evasão realizados por iniciativas de pesquisas de departamentos.

O objetivo do estudo realizado por Santos e Valverde [28] foi avaliar as mudanças que a reforma curricular de 1997 causou nos índices de evasão do departamento de Química. Estudos anteriores foram utilizados como parâmetros de comparação da evolução do problema ao longo do tempo, permitindo verificar se o impacto desse novo currículo foi negativo ou positivo e quantificá-lo.

O período de estudo determinado foi de 2000 a 2005, desconsiderando os três anos iniciais após a implementação do novo currículo. Tal medida foi tomada por tratarem-se de anos de maturação da nova grade, eliminando os chamados alunos híbridos, estudantes que desempenhavam as atividades do antigo currículo, mesmo com o novo em vigor, por ingressarem antes da reforma feita, o que poderia causar enviesamento nos resultados.

Os dados foram concedidos pelo Sistema de Graduação da Secretaria de Administração Acadêmica e da Secretaria de Planejamento da Universidade de Brasília. Também foram realizados estudos gerais de entrada e saída nos cursos vigentes dentro do departamento, formando, com isso, tabelas sobre o ingresso, a formatura e a evasão no período. Com essa análise geral, foi possível detectar quais foram os principais tipos de evasão juntamente com os períodos dos cursos mais propícios à evasão.

Tais informações gerais foram de suma importância para a delimitação das próximas análises. Como foi encontrado que, na maioria dos casos, os alunos se afastavam dos estudos nos quatro primeiros semestres e por motivos de não cumprimento de condição, foi decidido realizar a análise de índices de reprovação das disciplinas recomendadas de serem cursadas nesses semestres. Com essa análise, as autoras conseguiram detectar quais as disciplinas que mais impactam na evasão dos alunos nesse período. Ao final da pesquisa, observaram a ocorrência de uma diminuição significativa dos alunos evadidos juntamente com a diminuição do número de reprovações das disciplinas, o que reforça a pesquisa de Morosini et al. [23], que afirma que o rendimento acadêmico e a motivação estão fortemente relacionadas entre si e são grandes influenciadores dos índices de evasão.

Nesse trabalho foram realizadas apenas comparações estatísticas simples, as autoras não fizeram regressões que assegurem e quantifiquem as relações das variáveis analisadas, diferentemente do trabalho realizado por Lima e Costa [20] dentro do Departamento de Estatística que avalia os cursos formadores dos docentes das principais matérias do vestibular.

Lima e Costa [20] afirmam que a regressão logística é uma importante técnica para a verificação de quão forte está relacionada sua resposta com as possíveis causas dessa resposta. A regressão logística é importante para o trabalho delas por conta do mapeamento de um perfil dos alunos desses cursos e um perfil dos alunos não evadidos e evadidos para,

assim, levantar as principais causas da evasão. Com a aplicação das regressões é possível obter quais são os fatores que mais se relacionam com a evasão e com a permanência.

Lima e Costa [20] definem a regressão logística como um caso especial dos modelos lineares generalizados, em que a variável resposta é categórica, ou seja, a variável que você deseja prever está dividida em categorias e as variáveis explicativas, aquelas usadas para explicar a variável resposta, podem ser tanto categóricas quanto quantitativas, assumindo valores numéricos em determinada escala.

Para o trabalho de Lima e Costa [20], a variável resposta pode ser categorizada em **não evadido** e **evadido**. Já as variáveis explicativas foram determinadas pelos seguintes dados: sexo, habilitação, se o aluno foi monitor ou não, curso, número de ingressos, semestres de permanência, posição no fluxo e percentual de disciplinas com reprovação no período em estudo, dados provenientes do SIGRA.

Com a finalidade de conseguir a relevância dos dados explicativos com os dados respostas, foram realizados testes cíclicos, nos quais eram aplicadas diferentes técnicas em diferentes ciclos. Após tais procedimentos, conseguiram inferir um perfil dos estudantes por meio das variáveis relevantes. Os dados que apresentaram maiores relações foram sexo, turno, área do conhecimento, monitoria, habilitação, porcentagem de disciplinas obrigatórias com reprovação e semestre de permanência. Note a recorrência da relevância do estudo de reprovações em disciplinas e participação na vida acadêmica, medida pela participação em monitorias, reafirmando as ideias motivacionais e de desempenho de Morosini et al. [23].

Couto e Dantas [8], por meio de técnicas de mineração de dados para identificar padrões dos alunos dos três cursos oferecidos pelo Departamento de Ciência da Computação da UnB, construíram uma base de aprendizado sobre os principais perfis desses alunos. Os autores afirmam existir nos cursos do Departamento uma divisão forte nos principais motivos de evasão entre homens e mulheres. Apesar de terem perfis semelhantes, inclusive com taxas de evasão semelhantes, o principal motivo de evasão dos homens é o não cumprimento de condição, ou seja, não cumpriu o mínimo de desempenho em um período estipulado pela Universidade, enquanto o das mulheres é abandono do curso. Isso mostra que o desempenho dos homens está muito atrelado à sua saída e a das mulheres com a falta de interesse pelo curso, o que levou os autores a realizarem uma classificação dos dados seguindo os padrões de gênero.

Ainda nesse trabalho, Couto e Dantas [8] realizaram análises estatísticas dos dados e, por meio delas, afirmaram que o índice de reprovação nos primeiros semestres em disciplinas oferecidas pelo Departamento de Matemática, Instituto de Física e Departamento de Ciência da Computação está fortemente ligado à taxa de evasão por possuírem as piores médias de aprovação entre os estudantes do Departamento. Dessa maneira, o estudo foi conduzido também levando em consideração a média dos alunos em tais disciplinas. A classificação do aluno levou em consideração, ainda, o semestre de permanência, forma de ingresso no Departamento, raça, idade, entre outros fatores menos relevantes.

4.3 Evasão no Brasil

O projeto da Universidade Federal de Santa Catarina, de Souza et al. [33], faz uso das técnicas de estudo de evasão mais comuns, tais como as entrevistas e questionários.

Apesar do uso das entrevistas, a autora complementa com um estudo simples do perfil dos alunos por meio dos dados da universidade.

Para obtenção dos resultados esperados, os autores realizaram uma prévia seleção no banco de dados da universidade dos estudantes considerados evadidos, tendo como análise os dados referentes a forma de egresso desse alunos, mencionados anteriormente. Esses alunos foram contatados com um questionário de quarenta e nove perguntas que buscava responder questões de temas acadêmico-institucionais, sócio-político-econômicos e de ordem pessoal.

Das 153 respostas do total de 456 questionários enviados por e-mail, a maioria estava alinhada com a vertente Sócio-Político-Econômicos, principalmente a necessidade de trabalhar dos estudantes, mas também foi constatado a falta de interesse dos alunos com as matérias lecionadas na universidade. Juntamente com a análise dos dados desses alunos evadidos, foi visto que a maioria desses alunos possuíam baixo rendimento universitário, tal rendimento neste trabalho levou em consideração as notas que o aluno retirou nas disciplinas que ele cursou.

Já a pesquisa realizada por Vitelli et al. [34] na Universidade do Vale do Rio dos Sinos faz uma abordagem também mais objetiva, realizando o estudo de evasão da universidade com os próprios dados que ela possui. O objetivo era indicar os fatores que ocasionavam a evasão e quantificar a relevância deles, ou seja, o quão forte cada um está relacionado com o tema em questão.

O período de recolhimento de dados a serem analisados foi de cinco anos. Esse período foi escolhido uma vez que, dentro dele, a maioria dos cursos analisados teriam seu início e conclusão. Com o período definido, foi pedido à instituição dados dos alunos que poderiam ser de certa relevância ao estudo em questão. São eles: sexo, idade, estado civil, local de residência, média de desempenho nas atividades, média de desempenho no vestibular, quantidade de atividades matriculadas, percentual de atividades reprovadas, percentual de atividades aprovadas, percentual de atividades canceladas, percentual de atividades sem frequência (desistência), percentual de créditos concluídos, inadimplentes, três semestres contínuos sem matrícula, média de créditos matriculados por semestre, ajuda financeira, transferência interna, forma de ingresso, área (curso) e tamanho do curso (em créditos).

Com esses dados, foram aplicados os métodos da regressão logística. A escolha dos métodos foi baseada em uma série de indicadores que compõem a regressão logística, os quais foram testados com várias aplicações nos dados obtidos. Os indicadores escolhidos dentro da regressão logística são descritos na Tabela 4.1.

Após inúmeros testes realizados, foi construído um modelo que pudesse prever as chances de um estudante vir a evadir ou não da universidade. Apesar dos resultados satisfatórios, Vitelli et al. [34] advertem que as escolhas das variáveis a serem analisadas influenciam muito no resultado final e recomendam que os testes sejam refinados de ano em ano para a atualização de uma nova equação devido a mudanças no resultado dependendo do período analisado. No modelo apresentado pelos autores, os fatores mais impactantes da evasão estão relacionados ao desempenho acadêmico e ao tempo de curso.

No trabalho de Campello e Lins [6], da Universidade Federal de Pernambuco, os autores também buscam obter previsões de evasão e retenção dos estudantes no curso de engenharia de produção analisando seus dados. Esse trabalho tenta prever dois tipos de fenômenos universitários sendo eles evasão, a que também se refere esse trabalho, e re-

Tabela 4.1: Tabela com as descrições e finalidade de cada indicador que compõe o método de regressão logística, segundo Vitelli et al. [34].

Indicadores	Descrição
Log Likelihood	Mostra a capacidade de o modelo estimar a probabilidade associada à ocorrência de determinado evento, quanto menor o valor deste indicador, maior o poder preditivo do modelo.
Cox & Snell R Square	Compara o desempenho de modelos concorrentes. Entre duas equações logísticas igualmente válidas, deve-se preferir a que apresente maior Cox & Snell R Square.
Hosmer e Lemeshow	É obtido através de um teste Qui-quadrado que consiste em dividir o número de observações em cerca de 10 classes e, em seguida, comparar as frequências preditas com as observadas. Caso exista diferença entre os valores, então o modelo não seria capaz de produzir estimativas e classificações muito confiáveis
Wald	Tem como objetivo verificar se cada parâmetro estimado é significamente diferente de zero. Essa estatística segue uma distribuição Qui-quadrado e quando a variável dependente tem um único grau de liberdade pode ser calculada elevando-se ao quadrado a razão entre o coeficiente que está sendo testado e o respectivo erro-padrão

tenção. A retenção se manifesta quando um aluno ultrapassa o tempo proposto para o curso, podendo esse concluir após dado período ou não.

Como metodologia, foram determinados também o período de entrada que seria analisado e os dados dos estudantes que poderiam estar considerados como retidos ou evadidos. Para a extração desses dados, foram utilizadas técnicas de mineração de dados. De acordo com Campello e Lins [6], foram seguidas as seguintes etapas em todo o processo:

- Preparação dos dados – Consiste nas seguintes atividades:
 - Seleção de variáveis – Seleção que ocorre por possibilidade de relevância de cada variável.
 - Seleção de registros – Estimação do período. Essa seleção é importante devido a utilização das técnicas de análise precisarem de muito processamento, restringindo os dados a serem analisados com a finalidade de se otimizar tal processamento.
 - Transformação de variáveis – Alteração de alguns dados para o formato contínuo, ou seja, dividido em faixas.
- Seleção dos modelos de mineração de dados – É preciso explorar as várias alternativas de modelos a fim de encontrar aquele que é mais adequado ao problema.

- Validação do modelo – Após a seleção e dimensionamento do modelo, deve-se avaliar e interpretar os resultados, observando a taxa de precisão do modelo definido.
- Monitoração do modelo – É preciso monitorar permanentemente o modelo em função das mudanças ambientais que afetam os resultados e ajustar de acordo com a necessidade.

Seguindo esses passos, na etapa de seleção dos processos de mineração de dados cada técnica possui diversos tipos de métodos estatísticos. Essas técnicas são descritas na Tabela 4.2. Dentre elas, o pesquisador deve escolher as que apresentam melhor encaixe com o problema.

Tabela 4.2: Métodos de mineração de dados e respectivas descrição e finalidades, segundo Campello e Lins [6].

Métodos	Descrição
Estudo de <i>clustering</i>	Processos com objetivo de agrupar objetos similares para simplificar a identificação dos dados.
Regras de associação	Processos descritivos que visam estudar as relações ou associações (<i>links</i>) entre os dados.
Árvores de decisão	É uma forma de representar um conjunto de regras que definem uma precedência entre classes ou valores.
Estudo da vizinhança	Técnicas de previsão que interpretam o conteúdo das instâncias como pontos em um “espaço de dados” e pode dizer quais as variáveis estão interligadas com quais.

Após a aplicação dos métodos escolhidos, o trabalho teve como resultado a classificação dos estudantes em excelentes, bons, regulares, fracos, péssimos e desinteressados. Em cada uma dessas classes, foram apresentados os principais fatores que poderiam influenciar os alunos a desistirem do curso, como: desempenho escolar, problemas do sistema da universidade, formação básica deficiente, etc. Também foram apresentadas as disciplinas que possuem maiores chances de influenciar um abandono devido à sua taxa de reprovação. Segundo Campello e Lins [6], essa abordagem deve ser feita de maneira única para cada um dos cursos devido às suas características e perfil, não podendo generalizar esses resultados para os demais cursos da universidade.

4.4 Comparações dos trabalhos

Com a finalidade de facilitar a visualização dos diferentes métodos utilizados nos trabalhos analisados, foram construídas duas tabelas comparativas. A Tabela 4.3 mostra os dados que foram utilizados nas análises de cada trabalho.

Para a melhor organização das tabelas, foi dado a cada trabalho o nome da universidade de origem dos autores e, em caso de mesma universidade, foi adicionado o curso de origem do autores.

4.4.1 Dados

Como foi dito, a Tabela 4.3 relata quais foram os dados brutos utilizados em cada um dos trabalhos categorizados nas seguintes classes:

- Desligamento: Dados de egresso de alunos da IES. O desligamento pode ocorrer voluntariamente ou por condicionamento da IES.
- Abandono: Dados de egresso de alunos que não realizaram a rematrícula e não formalizaram o pedido de desvínculo com a IES.
- Transferência de curso: Dados de alunos que solicitam a alteração do curso para outro da mesma IES.
- Transferência de IES: Dados de alunos que solicitam a transferência de seus estudos para as dependências de uma nova IES.
- Pessoais: Dados de caráter pessoal que foram fornecidos no ato de matrícula, sendo, assim, produzidos fora do âmbito da IES, como sexo, endereço, escola de origem, etc.
- Evolução acadêmica: Dados acadêmicos que possuem a temporalidade como fator importante e que compõem o rendimento desse estudante dentro da IES categorizados dentro dos semestres vivenciados. Esses dados são relacionados com disciplinas matriculadas, menções, créditos aprovados, créditos reprovados, etc.
- Acadêmicos: Dados acadêmicos que não possuem a temporalidade com fator importante, como índice de rendimento acadêmico, porcentagem de créditos realizados, etc.
- Grau de Análise: Nível de requerimento hierárquico dos dados que foram analisados, se a análise foi em âmbito de IES ou de curso.

Na Tabela 4.3 é representado por um X o projeto que continha, em suas análises, os dados categorizados na classe representante da respectiva linha da tabela.

4.4.2 Metodologias

Foi realizado um resumo geral das decisões dos autores no decorrer da aplicação das metodologias presentes na revisão sistemática de cada trabalho. Com essa revisão se teve com maior clareza de quais eram os passos a serem seguidos nessa trabalho, além de obter dados para uma futura comparação de resultados. Além disso o processo de revisão dos trabalhos foi importante como indicativo da possibilidade de estudo de evasão através de uma metodologia objetiva como mineração e análise de dados. Dessa maneira para melhor a visualização do método percorrido pelo autores e de suas as decisões em cada uma delas, foram classificadas e resumidas as seguintes etapas.

- Objetivo: objetivos finais apresentados nos trabalhos analisados que compuseram o entendimento do fenômeno de evasão estudado, apresentados na Tabela 4.4.

Tabela 4.3: Categorias de dados que foram utilizados nos métodos estatísticos dos trabalhos analisados por esse projeto.

Tipo de dado	Granada [3]	Sinos [34]	UFSC [33]	Matérias básicas / UnB [20]	Química UnB [28]	Eng. de Produção / UFPE [6]	UFPE [31]
Desligamento			X	X	X	X	X
Não matrícula	X	X	X	X	X	X	X
Transferência de curso		X	X	X		X	X
Transferência de IES		X	X	X		X	X
Pessoais	X	X		X		X	X
Evolução acadêmica					X	X	
Acadêmicos	X	X	X	X	X	X	X
Instituições analisadas	Três faculdades	IES Privada	IES Pública	Cursos básicos do vestibular	Curso	Curso	IES Pública

Tabela 4.4: Objetivos apresentados nos trabalhos analisados que compuseram o entendimento do fenômeno de evasão estudado.

Estudo	Objetivos
Granada [3]	Perfil do aluno e principais fatores de saída
Sinos [34]	Perfil do aluno
UFSC [33]	Principais fatores de saída
Matérias básicas UnB [20]	Perfil dos alunos e principais fatores de saída
Química UnB [28]	Perfil do aluno e avaliação do novo PPP
Eng. de Produção UFPE [6]	Perfil do aluno e Principais fatores de saída
UFPE [31]	Perfil do aluno retido

- Extração dos Dados: formas de realização da extração dos dados, apresentadas na Tabela 4.5. Algumas possibilidades são análise de dados, mineração de dados e questionários.
- Métodos Estatísticos: métodos estatísticos empregados, apresentados na Tabela 4.6,

Tabela 4.5: Formas de realização da extração dos dados apresentadas nos trabalhos analisados que compuseram o entendimento do fenômeno de evasão estudado.

Estudo	Extração dos dados
Granada [3]	Mineração de dados
Sinos [34]	Mineração de dados
UFSC [33]	Entrevista de questionários e análise de Banco de dados
Matérias básicas UnB [20]	Análise de Banco de dados
Química UnB [28]	Análise de Banco de dados
Eng. de Produção UFPE [6]	Mineração de dados
UFPE [31]	Mineração de dados CRISP

para entender os dados e, a partir deles, retirar as informações necessárias para o entendimento da evasão.

Tabela 4.6: Métodos estatísticos apresentados nos trabalhos analisados que compuseram o entendimento do fenômeno de evasão estudado.

Estudo	Métodos estatísticos
Granada [3]	Regressão estatística
Sinos [34]	Regressão logística
UFSC [33]	Tabulação e análise estatística
Matérias básicas UnB [20]	Regressão logística múltipla e simples
Química UnB [28]	Tabulação e análise estatística
Eng. de Produção UFPE [6]	Geração de histogramas e regressão logística
UFPE [31]	Regressão estatística

- Período: duração dos períodos em que as análises foram realizadas, apresentadas na Tabela 4.7.
- Descarte de dados: dados que tiveram que ser descartados, apresentados na Tabela 4.8, para que os métodos apresentassem melhor rendimento ou acurácia em seus resultados
- Tratamento dos dados: após a aquisição dos dados, quais foram as medidas realizadas com esses, se houve algum tratamento antes da realização do processo de análise e quais foram esses procedimentos. Esses procedimentos são apresentados na Tabela 4.9.

Tabela 4.7: Períodos analisados que compuseram o entendimento do fenômeno de evasão estudado.

Estudo	Período
Granada [3]	15 anos
Sinos [34]	—
UFSC [33]	2 anos
Matérias básicas UnB [20]	—
Química UnB [28]	5 anos
Eng. de Produção UFPE [6]	6 anos
UFPE [31]	10 anos

Tabela 4.8: Dados que foram descartados apresentados nos trabalhos analisados que compuseram o entendimento do fenômeno de evasão estudado.

Estudo	Descarte de dados
Granada [3]	Não
Sinos [34]	—
UFSC [33]	Não
Matérias básicas UnB [20]	Retirada dos alunos com dados incompletos
Química UnB [28]	Alunos de transferência
Eng. de Produção UFPE [6]	Alunos que iniciaram antes do período determinado e saída de alunos nos 2 primeiros semestres
UFPE [31]	Alunos que iniciaram antes do período determinado, alunos que concluíram dentro do período determinado e dados discrepantes

- Novas Variáveis: variáveis, apresentadas na Tabela 4.10, criadas a partir dos dados brutos antes da realização dos procedimentos de análise. Nela ocorre a criação de faixas de precisão, variáveis Dummy, que são, segundo Campello e Lins [6], variáveis de valor binário que representam a presença ou não de uma característica, e cálculos de médias a partir de uma representação de dados temporais.

Tabela 4.9: Procedimentos de tratamento de dados apresentados nos trabalhos analisados que compuseram o entendimento do fenômeno de evasão estudado.

Estudo	Tratamento dos dados
Granada [3]	Preenchimento de dados incompletos e junção dos dados semestrais
Sinos [34]	Não
UFSC [33]	—
Matérias básicas UnB [20]	Não
Química UnB [28]	—
Eng. de Produção UFPE [6]	Identificação dos alunos entendidos como evadidos ou retidos e classificação dos alunos por rendimento
UFPE [31]	Dados incompletos tratados

Tabela 4.10: Novas variáveis criadas nos trabalhos analisados que compuseram o entendimento do fenômeno de evasão estudado.

Estudo	Novas variáveis
Granada [3]	Variáveis Dummy e variáveis em função do tempo (médias)
Sinos [34]	Variáveis Dummy
UFSC [33]	Variáveis de satisfação
Matérias básicas UnB [20]	Variáveis em função do tempo (médias)
Química UnB [28]	Não
Eng. de Produção UFPE [6]	Variáveis Dummy
UFPE [31]	Variáveis em função do tempo e variáveis calculadas a partir de outros dados

Capítulo 5

Análise de Dados no BCC - UnB

Para uma melhor organização do projeto realizado, ele foi subdividido em duas partes. A primeira, apresentada nesse capítulo, é a de análise de dados e compreende desde as primeiras decisões de projeto até a análise da situação do curso por meio de consultas nos dados obtidos. A segunda é a de mineração de dados, detalhada no capítulo seguinte.

Nesse capítulo, a Seção 5.1 apresenta como será tratada a evasão nesse trabalho. Na Seção 5.2 se encontram os dados obtidos e como foi a solicitação deles e, na Seção 5.3, o tratamento feito nos dados para deixar no formato correto para a análise e mineração de dados. A Seção 5.4 apresenta as perguntas definidas para guiar a análise, que é apresentada na Seção 5.5.

5.1 Evasão para esse trabalho

A primeira etapa do projeto, conforme explicado por Lobo [21], foi definir o que seria considerado por evasão. Optamos pela evasão de curso, ou seja, situação em que um aluno, estando matriculado em um determinado curso de uma determinada instituição de ensino superior, solicita transferência de curso ou IES, é desligado por vontade própria ou por exclusão da IES.

O curso escolhido para análise foi o Bacharelado em Ciência da Computação da Universidade de Brasília (BCC - UnB).

Serão considerados como evadidos os estudantes com tipos de evasão definidos de acordo com o regulamento da Universidade de Brasília [9] sobre tal conceito:

1. Abandono – Forma de exclusão do cadastro discente da UnB aplicada ao aluno que, durante um período letivo, não tenha efetivado matrícula em disciplinas, ou que, embora matriculado, tenha sido reprovado com SR (sem rendimento) em todas as disciplinas.
2. Desligamento por Jubilamento – Forma de exclusão do cadastro discente da UnB aplicada ao aluno que esgotou o tempo máximo de permanência previsto pelo Conselho Nacional de Educação – CNE para conclusão do curso.
3. Transferência para outra instituição de ensino superior – Forma de exclusão do cadastro discente da UnB aplicada ao aluno que, por iniciativa própria, mediante solicitação formal e apresentação de declaração de reserva de vaga, tenha assegurada

a sua admissão por transferência facultativa ou obrigatória em outra instituição de ensino superior, nacional ou estrangeira, para continuação de estudos.

4. Transferência interna – Saída do curso por meio de transferência para outro curso da própria Universidade.
5. Desligamento por não-cumprimento de condição – Forma de exclusão do cadastro discente da UnB aplicada ao aluno que, tendo sido anteriormente identificado como provável desligado por rendimento acadêmico ou por tempo de permanência, não tenha cumprido, no decorrer do(s) período(s) fixado(s), a condição que lhe foi imposta pelos órgãos colegiados.
6. Desligamento Voluntário – Forma de exclusão do cadastro discente da UnB aplicada ao aluno que, por iniciativa própria, tenha desistido de seu vínculo com a Universidade em determinado curso.

5.2 Aquisição dos dados

Os dados foram solicitados ao Decanato de Planejamento e Orçamento (DPO), que recuperou os dados do SIGRA - Sistema de Informação Acadêmica de Graduação. Foram solicitados tantos dados pessoais quanto acadêmicos dos estudantes do Bacharelado em Ciência da Computação.

Os dados obtidos vieram divididos em duas planilhas no formato do Excel. Uma delas continha os dados dos 2034 alunos matriculados desde a criação do curso até o 1º semestre de 2014. A Tabela 5.1 apresenta os dados que estavam presentes nela, juntamente com uma breve explicação sobre eles.

Tabela 5.1: Variáveis relacionadas aos estudantes do Bacharelado em Ciência da Computação.

Variável	Descrição
MatricAluno	Código do aluno no sistema.
CodCurso	Código do curso.
CodOpcao	Código da opção.
PrioridadeOpcao	Prioridade da opção, que pode ser primária ou secundária.
AnoIngresso	Ano de ingresso na Universidade.
SemestreIngresso	Semestre de ingresso na Universidade.
FormaIngresso	Forma de ingresso na Universidade.
AnoSaida	Ano de saída da Universidade.
SemestreSaida	Semestre de saída da Universidade.
FormaSaida	Forma de saída da Universidade.
PerIngressoOpcao	Período de ingresso na opção.
SemestreIngressoOpcao	Semestre de ingresso na opção.
ForIngressoOpcao	Forma de ingresso na opção.

Continua na página seguinte

Tabela 5.1 – *Continuação da tabela da página anterior*

Variável	Descrição
AnoIngressoOpcao	Ano de ingresso na opção.
MesIngressoOpcao	Mês de ingresso na opção.
PerSaidaOpcao	Período de saída da opção.
SemestreSaidaOpcao	Semestre de saída da opção.
ForSaidaOpcao	Forma de saída da opção.
AlunoRegistrado	Se aluno está registrado ou não.
PeriodoCurricular	Ano do período curricular.
SemestrePeriodoCurricular	Semestre do período curricular.
NumInscricaoCurricular	Número de inscrição no período curricular.
Judicial	Se aluno possui ou não algum processo judicial.
PoloAluno	Pólo do aluno.
IRA	Índice de Rendimento Acadêmico do aluno.
PeriodoProvavelFormando	Período em que o aluno será considerado como provável formando, segundo as regras da Universidade.
Obs	Campo aberto para observações sobre o aluno.
AluSexo	Sexo do aluno.
AluNivel	Nível de formação do curso.
AluNacionalidade	País de nascimento do aluno.
AluDtNasc	Data de nascimento do aluno.
AluPaisNasc	País de nascimento dos pais.
AluUfNasc	UF de nascimento do aluno.
AluCotId	Qual o sistema pelo qual o aluno entrou, que pode ser universal ou por alguns dos tipos de cotas aceitas pela Universidade.
AluPne	Portador de necessidades especiais.
AluRA	Registro acadêmico do aluno (matrícula).
AluEscola	Tipo da escola que o aluno se formou no ensino médio, que pode ser privada ou pública.
Alucodigoracacor	Cor do aluno, declarada por ele.
CampoNaoInd	Cidade de nascimento do aluno.
AluPesCodigoPessoa	Código de cadastro do aluno.
OpcDenominacao_last	Nome da opção.
OpcIni_last	Período de início da opção na Universidade.
OpcPerFim_last	Período de fim da opção na Universidade.
OpcCredFormat_last	Quantidade de créditos necessários para se formar no curso.
CurNivel_last	Nível do curso.
OpcMinPerm_last	Permanência mínima no curso, em semestres.
OpcMaxPerm_last	Permanência máxima no curso, em semestres.
OpcGrau_last	Grau da opção.
OpcDuracao_last	Duração estimada da opção.
OpcDataResol_last	Data de resolução da criação da opção.
OpcHorasAtivExtensao_last	Horas exigidas de atividades de extensão.

Continua na página seguinte

Tabela 5.1 – *Continuação da tabela da página anterior*

Variável	Descrição
CurPerIni_last	Período em que o curso foi criado.
CurPerFim_last	Período em que o curso deixou de ser oferecido.
CurDenominacao_last	Nome do curso.
CurOrgResp_last	Órgão responsável pelo curso.
CurDepto_last	Departamento que oferece o curso.
CurAreaCon_last	Área de conhecimento do curso.
CurDuracao_last	Duração estimada do curso.
CurTurno_last	Turno em que o curso é oferecido.
CurForma_last	Forma como o curso é oferecido.
CurCodigoEnade_last	Código do ENADE para o curso.
Cód.ENADE	Código do ENADE para o curso.
Grau	Grau do curso.
Modalidade	Modalidade do curso.
Nome	Nome do curso.
Turno	Turno do curso.
LocaldeOferta	<i>Campus</i> em que o curso é oferecido.
NomeAnuário	Nome do curso no anuário.
UnidadeCód	Código da unidade do departamento que oferece o curso.
Unidade	Nome da unidade do departamento que oferece o curso.
DeptoCód	Código do departamento que oferece o curso.
Departamento	Nome do departamento que oferece o curso.

A segunda planilha continha os dados referentes ao histórico de todos os matriculados no curso de 2000 até 2013. Ao todo, foram 52900 registros. A Tabela 5.2 mostra quais os dados relacionados ao histórico.

Tabela 5.2: Variáveis relacionadas ao histórico dos estudantes já matriculados no Bacharelado em Ciência da Computação.

Variável	Descrição
MatricAluno	Código do aluno no sistema.
Ano	Ano em que a disciplina foi cursada.
Semestre	Semestre do ano curricular em que a matéria foi cursada.
Disciplina	Código da disciplina cursada.
Turma	Código da turma da disciplina cursada.
Menção	Menção obtida pelo aluno após a conclusão da disciplina cursada.
Percentual de faltas	Quantidade de faltas, em porcentagem, do aluno na disciplina cursada.

5.3 Tratamento dos dados

Com intuito de facilitar a manipulação dos dados, foi criado um banco de dados MySQL com as informações das planilhas. Cada uma das planilhas se tornou uma tabela do banco de dados criado. Para isso, primeiramente a planilha foi convertida para o formato CSV, o que permitiu a utilização do *CSV to SQL Converter*¹. Essa ferramenta permite exportar um arquivo CSV e gera um arquivo SQL com base nele. Em seguida, esse arquivo foi exportado para o phpMyAdmin. Os dados da Tabela 5.1 se tornaram a tabela **alunos** do nosso banco, enquanto os da Tabela 5.2 se tornaram a tabela **historico**.

Com o banco criado, foi possível iniciar o tratamento dos dados, que corresponde às primeiras etapas do processo de extração de conhecimento, ou seja, limpeza, integração, seleção e transformação dos dados. A consulta SQL necessária para realização dessa etapa de tratamento de dados está disponível no Apêndice A.

Limpeza de dados

Durante a limpeza de dados, foram removidas as seguintes variáveis, anteriormente descritas na Tabela 5.1: CodCurso, CodOpcao, PrioridadeOpcao, MesIngressoOpcao, AlunoRegistrado, Judicial, PoloAluno, IRA, PeriodoProvavelFormando, Obs, AluNivel, AluRA, CampoNaoInd, AluPesCodigoPessoa, OpcDenominacao_last, OpcIni_last, OpcPerFim_last, OpcCredFormat_last, CurNivel_last, OpcMinPerm_last, OpcGrau_last, OpcMaxPerm_last, OpcDuracao_last, OpcDataResol_last, OpcHorasAtivExtensao_last, CurPerIni_last, CurPerFim_last, CurDenominacao_last, CurOrgRest_last, CurAreaCon_last, CurDuracao_last, CurTurno_last, CurForma_last, CurCodigoEnade_last, CurDepto_last, Cód. Enade, Grau, Modalidade, Nome, Turno, LocaldeOferta, Nome-Anuario, UnidadeCód, Unidade, Unidade, DeptoCód e Departamento.

Integração de dados

A integração foi feita relacionando os dados do aluno com seu histórico. Para isso, foi feito um *INNER JOIN* no atributo MatricAluno das duas tabelas. Além disso, foi feita uma integração do histórico com os dados das matérias obrigatórias e optativas do curso. Para isso, foram criadas duas tabelas, uma para cada tipo de matéria, com informações do código da disciplina, nome da disciplina, departamento e, para as matérias obrigatórias, o semestre que é recomendado no fluxo.

Seleção de dados

Durante a seleção dos dados necessários para a aplicação, outros foram removidos. Primeiramente foram removidos os dados dos alunos que entraram antes de 2000 ou depois de 2013, pois não foram obtidos os históricos deles. Também foram retirados todos os dados de alunos, tanto pessoais quanto de histórico, que possuem forma de saída com códigos 9 e 55, segundo a Tabela 5.3, que apresenta os possíveis motivos de saída de uma pessoa juntamente com seu código.

¹<http://www.convertcsv.com/csv-to-sql.htm>

Tabela 5.3: Código e descrição das possíveis formas de saída de um aluno.

Código	Forma de saída
0	Cursando
1	Formatura
2	Desligamento - Rendimento acadêmico
3	Desligamento - Jubilamento
4	Desligamento - Falta de documentação
5	Desligamento - Força de convênio
6	Transferência
7	Desligamento - Voluntário
8	Expulsão disciplinar
9	Falecimento
10	Ex-aluno (Decreto 477)
11	Desligamento - Falta de pagamento
12	Desligamento - Decisão judicial
13	Formatura com especialização
14	Conclusão de estágio da pós-graduação
15	Formatura anterior a 1/88
16	Desligamento - Abandono
17	Desligamento - Não cumpriu condição
18	Desligamento com especialização
19	Formatura da pós-graduação
20	Reprovado 3 vezes na mesma disciplina obrigatória
21	Novo vestibular
22	Mudança de turno
50	Vestibular para outra habilitação
51	Vestibular para mesma habilitação
52	Mudança de curso
53	Mudança de habilitação
54	Confirmação de habilitação
55	Anulação de registro
99	Outros

Transformação de dados

Para deixar os dados na forma adequada para a aplicação foram feitas diversas transformações.

Foi criado um novo campo na tabela de alunos com a forma de saída dele podendo assumir quatro valores: 0 para cursando, 1 para formado, 2 para evadido ou 3 para outros casos. Com base nos códigos apresentados na Tabela 5.3, os alunos considerados como cursando eram apenas os que tinham forma de saída 0; os formados aqueles com código 1, 13, 14, 15 e 19; os evadidos com 2, 3, 4, 5, 6, 7, 8, 11, 12, 16, 17, 18, 20, 21, 50, 52 e 53; e os outros com valores 10, 22, 51, 54 e 99.

Foi adicionado, nessa mesma tabela, um campo para idade em que o aluno saiu do

curso e, caso ele ainda esteja cursando, esse campo recebe valor 0. Para realizar essa conta da idade, foi necessário alterar o formato da data de nascimento que já estava no banco para o formato DD/MM/YYYY.

Na tabela de histórico foi criado um campo chamado aprovação que pode assumir valor 0 para quando o aluno é reprovado e 1 para quando é aprovado. O aluno é reprovado quando obtém menção SR, II, MI, TJ ou TR e aprovado com menções MM, MS, SS ou CC. Os registros de alunos com menções AP ou DP não serão considerados pela falta de informação sobre o significado delas.

Ainda na tabela do histórico, foi adicionado um campo de tipo da disciplina, que pode assumir valor 0 para módulo livre, 1 para obrigatória e 2 para optativa.

Além desses campos, também foi adicionado um do semestre em que o aluno estava quando cursou aquela disciplina. Para calcular esse valor foi preciso verificar o semestre em que a matéria foi cursada e o semestre que ele entrou no curso. A Tabela 5.4 apresenta as fórmulas utilizadas para cálculo do semestre curricular em que ela foi cursada.

Tabela 5.4: Fórmulas utilizadas para cálculo do semestre curricular em que uma disciplina foi cursada em função do semestre em que ela foi feita e do que o aluno entrou no curso. A coluna semestre_entrada corresponde ao semestre de entrada do aluno em algum ano e a semestre_cursada corresponde ao semestre de algum ano em que a disciplina foi cursada. Nas fórmulas, a variável ano_entrada é o ano de entrada do aluno no curso e ano_cursada é o ano em que a disciplina foi cursada.

semestre_entrada	semestre_cursada	Fórmula
1	1	$2 * (ano_entrada - ano_cursada) + 1$
1	2	$2 * (ano_entrada - ano_cursada)$
2	1	$2 * (ano_entrada - ano_cursada) + 2$
2	2	$2 * (ano_entrada - ano_cursada) + 1$

Na tabela de alunos foram criados dois novos campos: desempenho médio e taxa de evasão. O desempenho médio é calculado pela quantidade de créditos realizados com aprovação pelo aluno dividido pela quantidade de créditos realizados. De forma semelhante, a taxa de reprovação é a quantidade de créditos com reprovação dividido pelo total de créditos cursados.

Ainda na tabela de alunos, por conta da diversidade de situações de alunos, por exemplo, alunos que entraram por transferência, que possuem histórico antes do ano que entraram ou então os que entraram diretamente no Bacharelado em Ciência da Computação, foi criado um campo de grupo de análise. Esse campo recebe valor 0 para quando o aluno entrou diretamente no curso e não possui histórico anterior, 1 para quando o aluno cursou outro curso na UnB antes de mudar para o Bacharelado e 2 para os casos de alunos que entraram diretamente no curso mas possuem histórico anterior. Nesse trabalho, apenas os casos do grupo de análise com valor 0 serão considerados.

5.4 Perguntas definidas

Antes de utilizar a técnica de mineração de dados, foi realizada uma análise dos dados para compreensão da situação do curso em relação à evasão.

Para guiar essa análise, foram definidas algumas perguntas que deveriam ser respondidas:

1. Qual o semestre que possui maior índice de evasão e qual o menor? Qual é o principal motivo de evasão em cada semestre?
2. Qual é o índice de evasão por sexo? E o de formatura? Quais os principais motivos de evasão por sexo?
3. Qual é o índice de evasão por idade? E o de formatura? Quais os principais motivos de evasão por idade?
4. Qual é o índice de evasão por tipo de escola, que pode ser pública ou particular? Quais os principais motivos de evasão por tipo de escola?
5. Qual é o índice de evasão por forma de ingresso? Qual é a forma que possui maior índice e qual a menor?
6. Qual é o desempenho médio, definido como quantidade de créditos com aprovação dividido pela quantidade de créditos realizados, dos alunos considerados como evadidos e dos que formaram?
7. Qual é a taxa de reprovação dos alunos considerados como evadidos e dos que formaram?
8. Qual é a média de créditos obrigatórios cursados por semestre dos alunos considerados como evadidos e dos que formaram?
9. Qual o índice de reprovação por departamento das matérias obrigatórias dos alunos evadidos e dos que formaram?
10. Quais as disciplinas obrigatórias, organizadas por semestre recomendado no fluxo, que possuem as maiores taxas de reprovação?
11. Qual é a variação na taxa de reprovação de cada disciplina obrigatória ao longo dos anos?
12. Qual o rendimento, ou seja, as menções, das disciplinas obrigatórias, organizadas por semestre recomendado no fluxo, dos alunos considerados como evadidos e dos que formaram?

5.5 Análise de dados

Com base nas perguntas definidas na seção anterior, foram criados os gráficos e as tabelas apresentados nessa seção. A consulta SQL utilizada para isso está disponível no Apêndice B. Todos os dados utilizados para criação dos gráficos apresentados nessa Seção estão no Apêndice C.

5.5.1 Semestre

A Tabela C.1 apresenta os dados da evasão por semestre e dos motivos de evasão em cada um deles. A Figura 5.1 apresenta o gráfico correspondente.

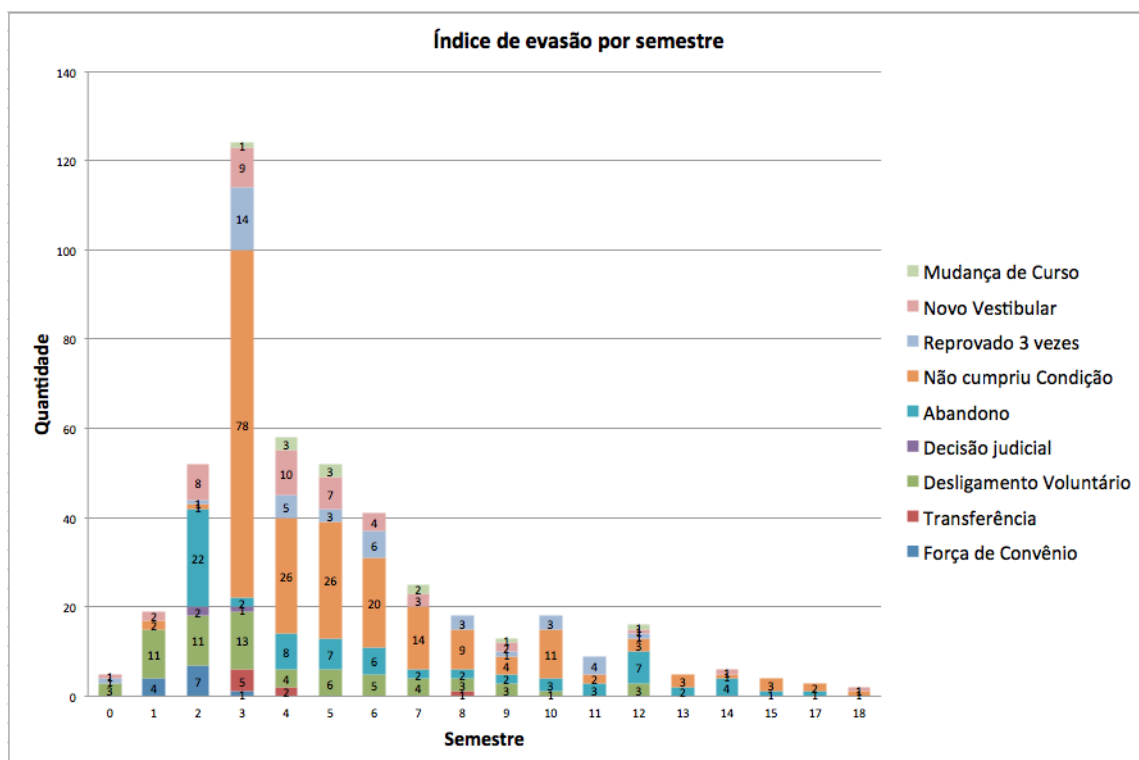


Figura 5.1: Taxa de evasão por semestre e motivos de evasão em cada um. O código do motivo de saída está de acordo com o apresentado na Tabela 5.3. O semestre 0 corresponde a um semestre cursado no verão, independente da posição do fluxo em que o aluno estava quando cursou. Cada uma das barras corresponde a um semestre, representado no eixo X. A altura da barra representa a quantidade de alunos que evadiram naquele semestre, de acordo com o eixo Y. As cores em cada uma das barras correspondem ao motivo de saída, segundo a legenda do lado direito do gráfico. O número dentro de cada uma das cores mostra a quantidade absoluta de alunos que evadiram naquele semestre e por qual motivo.

A Tabela C.2 apresenta os dados de formatura por semestre. A Figura 5.2 apresenta o gráfico correspondente.

O terceiro semestre é o que ocorre a maior taxa de evasão, que possui como principal motivo de saída o desligamento por não cumprimento de condição. Esse é o principal motivo de saída da maioria dos semestres. No caso do segundo semestre, que também possui elevado índice de evasão, o principal motivo é o desligamento por abandono, motivo que se repete como principal em outros semestres. Outros motivos de saída que são os principais em algum dos semestres são: desligamento voluntário e reprovar três vezes a mesma matéria obrigatória. Os semestres que mais alunos se formam são o 10º, 11º e 12º, sendo que o recomendado pelo fluxo são 9 semestres.

5.5.2 Sexo

A Tabela 5.5 apresenta os dados da quantidade de alunos já matriculados no curso separados por sexo e a situação deles, ou seja, se estão cursando, formados ou evadidos.

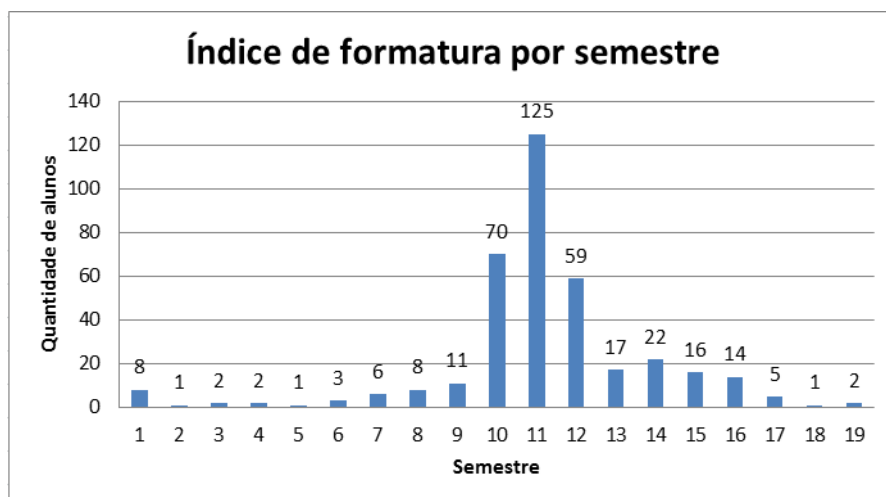


Figura 5.2: Taxa de formatura por semestre. Cada uma das barras corresponde a um semestre, representado no eixo X. A altura da barra representa a quantidade de alunos que evadiram naquele semestre, de acordo com o eixo Y.

Tabela 5.5: Quantidade de alunos que já saíram do curso separados por sexo e situação em que eles saíram.

Sexo	Feminino		Masculino	
	Quantidade	%	Quantidade	%
Formado	45	46.4	328	44.0
Evadido	52	53.6	418	56.0
Total	97	100	746	100

Ainda com base nessa separação por sexo, são analisadas as taxas de evasão e de formatura por sexo e os motivos de saída de quem evadiu. A Tabela C.3 mostra os dados femininos e a Figura 5.3 o gráfico correspondente.

Da mesma forma, a Tabela C.4 e a Figura 5.4 mostram os dados masculinos.

Apesar da quantidade absoluta de formados e evadidos do sexo masculino ser superior a do feminino, a porcentagem não possui uma diferença significativa. A taxa de evasão feminina é de 53.6%, enquanto a masculina é de 56.0%. Já a de formatura feminina é de 46.4% e a masculina de 44.0%. No entanto, existe uma diferença entre os principais motivos de evasão entre os dois sexos. As mulheres evadem por três principais motivos: desligamento por não cumprimento de condição, desligamento voluntário e desligamento por abandono. Já os homens evadem predominantemente por desligamento por não cumprimento de condição. Hoje existem no curso 38 mulheres e 287 homens.

5.5.3 Idade

A Tabela C.5 apresenta os dados da quantidade de alunos evadidos separados por idade e os motivos de saída deles. A Figura 5.5 apresenta o gráfico correspondente.

Da mesma forma, a Tabela C.6 e a Figura 5.6 mostram os dados de formatura separados por idade.

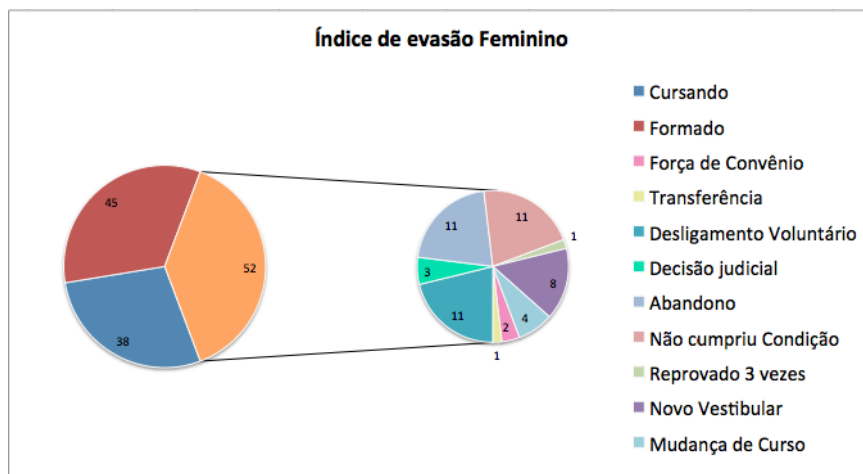


Figura 5.3: Taxa de evasão e de formatura feminina e motivos de saída de quem evadiu. O código do motivo de saída está de acordo com o apresentado na Tabela 5.3. O gráfico do lado esquerdo mostra a quantidade de evadidos, de formados e de alunos cursando, sendo o azul a quantidade de alunos cursando, o vermelho a de formados e o laranja a de evadidos. Já o do lado direito, mostra os motivos de saída dos evadidos, sendo os números no gráfico a quantidade de alunos que saíram por cada motivo.

A maioria dos alunos evadidos saíram com 20 anos, e a principal faixa de evasão de 19 a 22 anos. A partir de 19 anos, o principal motivo de evasão é de desligamento por não cumprimento de condição. O principal motivo de evasão aos 18 anos é de desligamento voluntário. Já aos 17 anos é desligamento por abandono.

Já dos formandos, a maioria tinha 23 anos quando saiu do curso. Essa idade de formatura ser a principal é justificada pela maioria dos alunos entrarem com 18 anos e permanecerem cinco anos no curso.

5.5.4 Tipo de escola

A Tabela C.7 apresenta os dados da quantidade de alunos já matriculados no curso separados por tipo de escola, que pode ser pública, particular ou não declarado. A Figura 5.7 apresenta o gráfico correspondente.

Dos alunos que declararam o tipo de escola, a maioria, estudou em escola particular. O principal motivo de saída dos alunos que evadiram, independente do tipo de escola, foi de desligamento por não cumprimento de condição. Como a maioria dos alunos, 51.54%, não declarou o tipo de escola, esses dados podem ser considerados enviesados.

5.5.5 Forma de ingresso

Os códigos referentes às possíveis formas de ingresso estão na Tabela 5.6. A Tabela C.8 apresenta os dados da quantidade de alunos já matriculados no curso separados por forma de ingresso. A Figura 5.8 apresenta o gráfico correspondente.

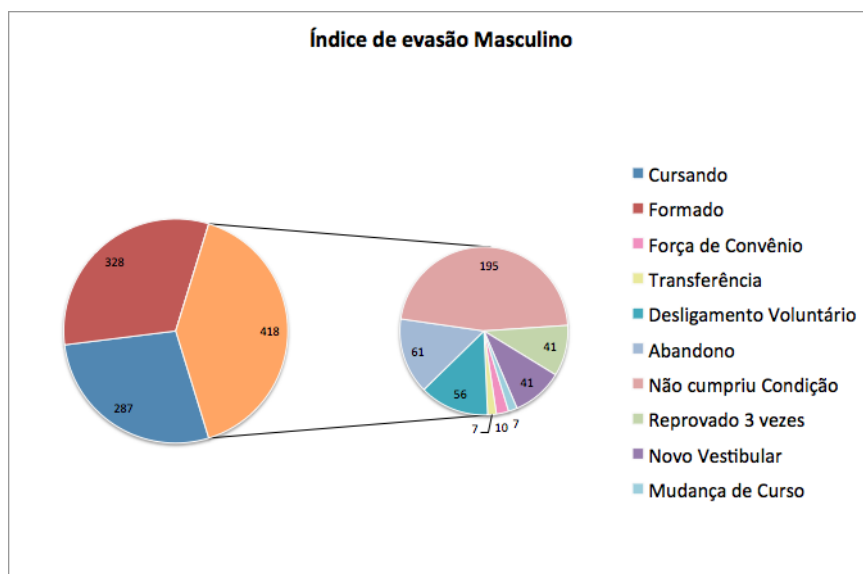


Figura 5.4: Taxa de evasão e de formatura masculina e motivos de saída de quem evadiu. O código do motivo de saída está de acordo com o apresentado na Tabela 5.3. O gráfico do lado esquerdo mostra a quantidade de evadidos, de formados e de alunos cursando, sendo o azul a quantidade de alunos cursando, o vermelho a de formados e o laranja a de evadidos. Já o do lado direito, mostra os motivos de saída dos evadidos, sendo os números no gráfico a quantidade de alunos que saíram por cada motivo.

Tabela 5.6: Código e descrição das possíveis formas de ingresso de um aluno.

Código	Forma de ingresso
1	Vestibular
2	Transferência Obrigatória
3	Transferência Facultativa
4	Portador Diplom Curso Superior
5	Acordo Cultural-PEC
6	Convênio-Int
7	Matrícula Cortesia
8	Seleção
9	Inscrição
10	Estágio
11	Convênio FEDF
12	Convênio UFPA
13	Novo Vestibular
14	Vestibular para mesmo Curso
15	Duplo Curso
16	Reintegração
17	Programa de Avaliação Seriada

Continua na página seguinte

Tabela 5.6 – *Continuação da tabela da página anterior*

Código	Forma de ingresso
18	PIE
19	Anistia
20	Convênio - Andifes
21	Intercâmbio de Pós-Graduação
22	Refugiado
23	Transferência-Convênio.
24	PEC-G Peppfol
25	Visitante
26	Convênio de Pós-Grad. PEC-PG
27	Enem
28	Mudança de Turno
29	Seleção para PARFOR
30	Convênio Funai
50	Dupla Habilitação
51	Mudança de Habilitação
52	Mudança de Curso
53	Registro de Habilitação
54	Registro de Diploma
55	Convênio UFMS
56	Seleção- Defesa Direta de Tese
57	Programa Bolsista PUC-GO
58	Convênio de Co-Tutela
99	Outros

Os alunos que mais evadem são os que entraram por algum tipo de convênio ou matrícula cortesia, chegando a ter uma taxa de evasão de 100%. A menor taxa de evasão é dos alunos que entraram por mudança de curso. As duas formas de ingresso mais comuns são por vestibular e pelo PAS. Dessas, a taxa de evasão do vestibular é superior a do PAS.

5.5.6 Desempenho médio

A Tabela C.9 apresenta os dados da quantidade de alunos evadidos separados por desempenho médio. A Figura 5.9 apresenta o gráfico correspondente.

Da mesma forma, a Tabela C.10 e a Figura 5.10 mostram os dados de formatura separados por desempenho médio.

O desempenho médio nas disciplinas obrigatórias diverge entre os evadidos e os formados. Nos evadidos percebe-se uma variação maior de valores de desempenho, possuindo estudantes com desempenhos em todas as possíveis faixas de valores. A maioria dos evadidos tinham desempenho médio inferior a 70%, ou seja, foram aprovados em menos de 70% dos créditos obrigatórios cursados. Já no caso dos formados, nenhum estudante teve desempenho inferior a 50% e a maioria conseguiu ter 90% de créditos cursados aprovados.

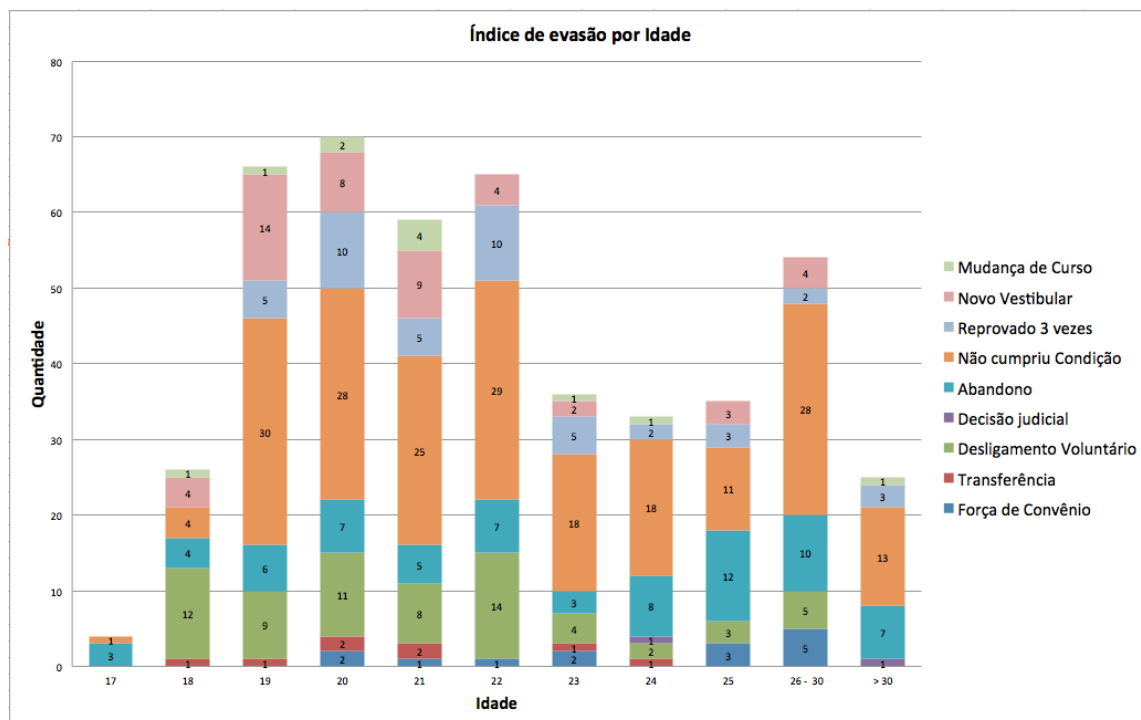


Figura 5.5: Índice de evasão por idade e por motivo de saída. O código do motivo de saída está de acordo com o apresentado na Tabela 5.3. Cada uma das barras corresponde a uma idade, representada no eixo X. A altura da barra representa a quantidade de alunos que evadiram com aquela idade, de acordo com o eixo Y. As cores em cada uma das barras correspondem ao motivo de saída, segundo a legenda do lado direito do gráfico. O número dentro de cada uma das cores mostra a quantidade absoluta de alunos que evadiram com aquela idade e por esse motivo.

5.5.7 Taxa de reprovação média

A Tabela C.11 apresenta os dados da quantidade de alunos evadidos separados por taxa de reprovação média. A Figura 5.11 apresenta o gráfico correspondente.

Da mesma forma, a Tabela C.12 e a Figura 5.12 mostram os dados de formatura separados por taxa de reprovação média.

A taxa de reprovação média nas disciplinas obrigatórias é calculada de forma que a soma do desempenho médio com a taxa de reprovação de cada aluno seja 1. Com isso, temos resultados semelhantes aos obtidos na análise do desempenho médio. Os evadidos possuem uma maior variação de taxa de reprovação média. Enquanto os evadidos reprovaram em mais de 30% dos créditos obrigatórios cursados, a maioria dos formados reprovaram menos de 30%.

5.5.8 Média de créditos obrigatórios por semestre

A Tabela C.13 apresenta os dados da média de créditos obrigatórios cursados por semestre dos evadidos. O semestre 0 representa um semestre cursado no verão, independente do período curricular que ele estava quando cursou. A Figura 5.13 apresenta o gráfico correspondente.

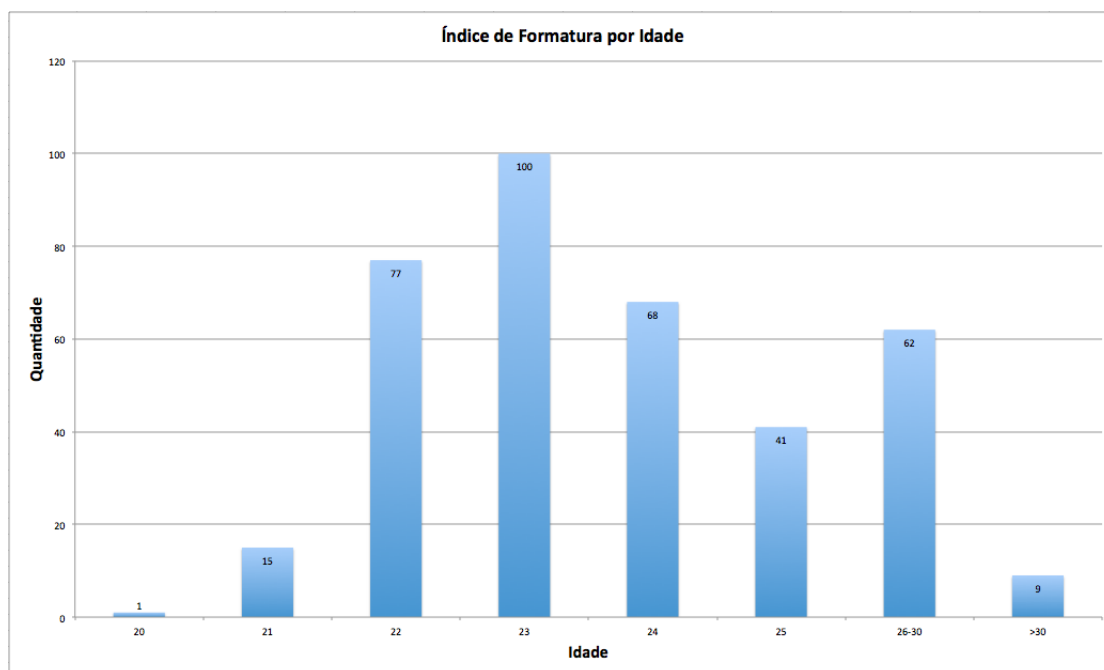


Figura 5.6: Índice de formatura por idade. Cada uma das barras corresponde a uma idade, representada no eixo X. A altura da barra representa a quantidade de alunos que evadiram com aquela idade, de acordo com o eixo Y.

Da mesma forma, a Tabela C.14 e a Figura 5.14 mostram as médias de créditos obrigatórios cursados por semestre pelos formados.

A média de créditos cursados por semestre dos formandos é superior ao dos evadidos até o sexto semestre. No sétimo semestre do fluxo é quando ocorre uma redução de matérias obrigatórias, de acordo com o currículo do BCC – UnB. É justamente a partir do sétimo semestre que ocorre uma diminuição na média de créditos obrigatórios cursados por semestre, ficando inferior à dos evadidos. A partir do décimo quarto semestre, que é o tempo máximo de permanência no curso, a situação volta a inverter.

5.5.9 Índice de reprovação por departamento

A Tabela C.15 apresenta os dados da taxa média de reprovação dos alunos evadidos e dos formados. A Figura 5.15 apresenta o gráfico correspondente.

As reprovações acontecem, principalmente, nas disciplinas oferecidas pelo Instituto de Física, Departamento de Matemática e Departamento de Ciência da Computação. Além disso, existe uma relação forte entre a taxa de reprovação nas disciplinas obrigatórias e a forma de saída do aluno.

5.5.10 Reprovações em disciplinas obrigatórias

As Tabelas C.16, C.17, C.18, C.19 e C.20 apresentam os dados da taxa de reprovação das disciplinas obrigatórias de acordo com o semestre do fluxo que são recomendadas. As Figuras 5.16, 5.17, 5.18, 5.19 e 5.20 apresentam os gráficos correspondentes. Foram selecionadas apenas disciplinas até o 5º semestre do fluxo e oferecidas pelo Departamento

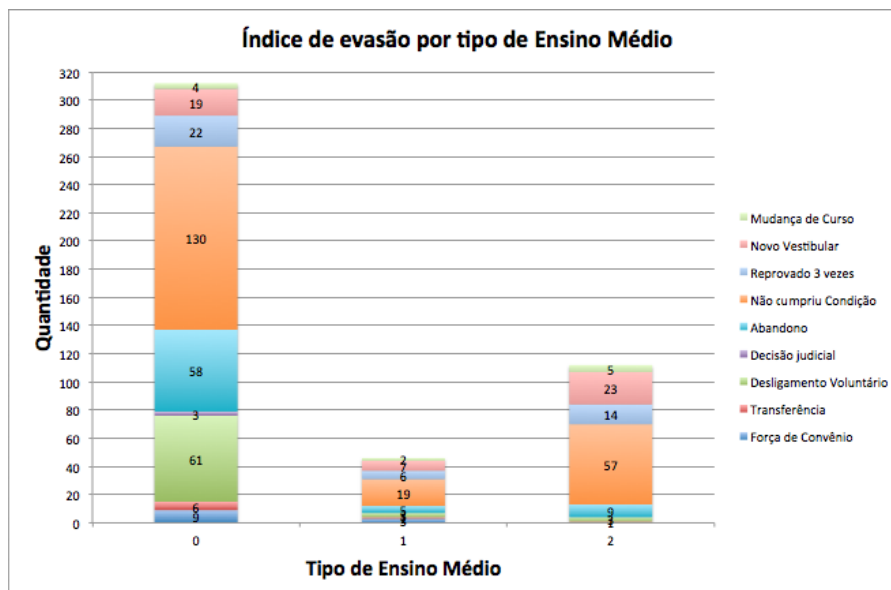


Figura 5.7: Taxa de evasão por tipo de escola e motivos de saída. O código do motivo de saída está de acordo com o apresentado na Tabela 5.3. Cada uma das barras corresponde a um tipo de escola, representado no eixo X. A altura da barra representa a quantidade de alunos que evadiram que possuem aquele tipo de escola, de acordo com o eixo Y. As cores em cada uma das barras correspondem ao motivo de saída, segundo a legenda do lado direito do gráfico. O número dentro de cada uma das cores mostra a quantidade absoluta de alunos que evadiram daquele tipo de escola e por esse motivo.

de Ciência da Computação, Departamento de Matemática ou Instituto de Física. A escolha do período ocorreu por incluir os semestres com maiores índices de evasão. Já a do departamento que oferece a disciplina foi devido ao fato de serem os que possuem disciplinas obrigatórias com maiores taxas de reprovação.

A análise dessas taxas de reprovação mostra que determinadas disciplinas são de difíceis aprovações. Essas disciplinas, de forma geral, são pré-requisito de outras, o que explica a retenção dos alunos no curso. Considerando uma alta taxa de reprovação valores maiores ou iguais a 25%, o primeiro semestre pode ser considerado crítico, uma vez que todas as disciplinas do fluxo desse semestre possuem alta taxa de reprovação. Como não existe separação na taxa de reprovação de evadidos e formados, não é possível determinar o quanto essa taxa de reprovação de cada disciplina está relacionada com a evasão no curso.

5.5.11 Taxa de reprovação ao longo dos anos

As Tabelas C.21, C.22, C.23, C.24 e C.25 apresentam os dados da variação da taxa de reprovação ao longo dos anos das disciplinas obrigatórias de acordo com o semestre do fluxo que são recomendadas. As Figuras 5.21, 5.22, 5.23, 5.24 e 5.25 apresentam os gráficos correspondentes. Foram selecionadas apenas disciplinas com alta taxa de reprovação, ou seja, com valores maiores ou iguais a 25%, de acordo com os resultados obtidos na Subseção 5.5.10.

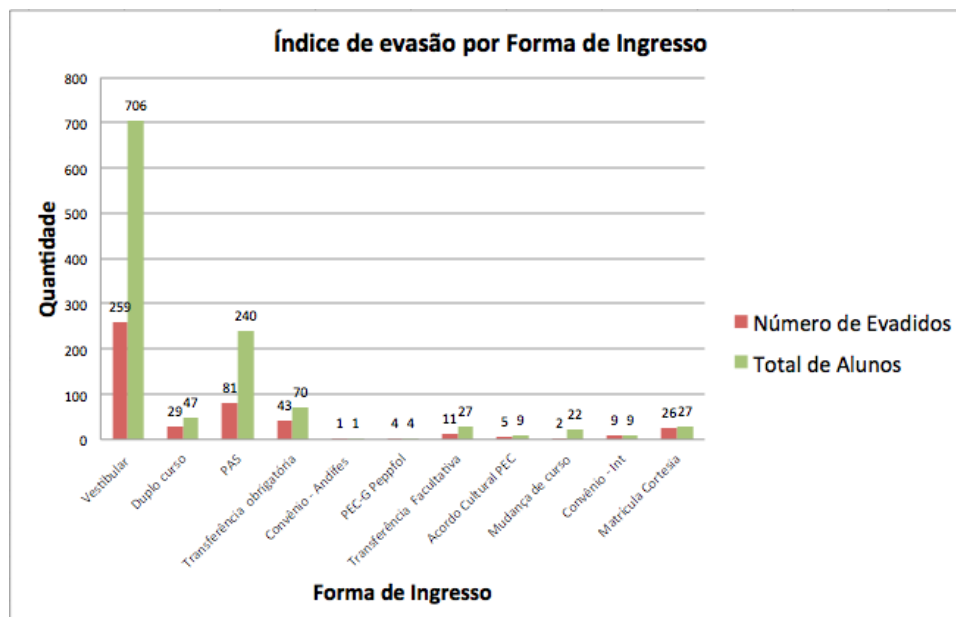


Figura 5.8: Taxa de evasão por forma de ingresso. O código da forma de ingresso está de acordo com o apresentado na Tabela 5.6. Cada uma das barras corresponde a uma forma de ingresso, representada no eixo X. A altura da barra representa a quantidade de alunos que evadiram que possuem aquele tipo de forma de ingresso, de acordo com o eixo Y. As barras vermelhas correspondem aos alunos que evadiram e a verde ao total que entraram por aquela forma de ingresso, de acordo com a legenda do lado direito do gráfico.

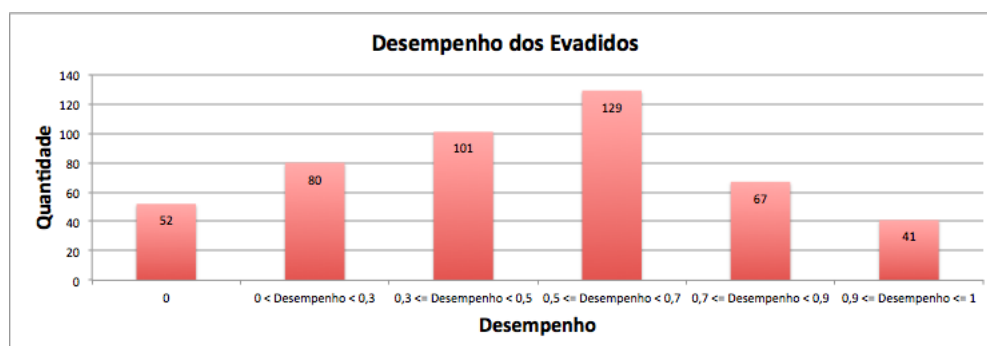


Figura 5.9: Desempenho médio dos evadidos organizados em faixas de desempenho. Cada uma das barras corresponde a uma faixa de desempenho, representada no eixo X. A altura da barra representa a quantidade de alunos evadidos que possuem desempenho dentro daquela faixa, de acordo com o eixo Y.

A análise da evolução das taxas de reprovações mostra um aumento geral na quantidade de reprovações a cada ano. De forma geral, as taxas de reprovações começaram a aumentar consideravelmente a partir de 2009. Considerando a quantidade média de semestres que um aluno leva para se formar e da quantidade média de créditos obrigatórios cursados por semestre, temos que grande parte dos dados desse período de aumento das taxas de reprovações são de alunos que estão cursando atualmente ou já evadiram.

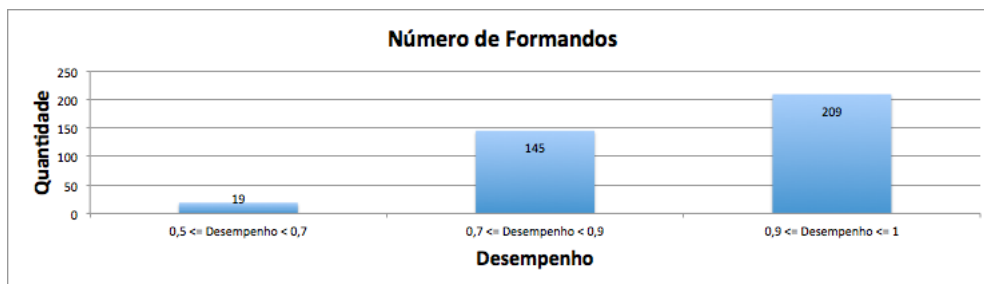


Figura 5.10: Desempenho médio dos formados organizados em faixas de desempenho. Cada uma das barras corresponde a uma faixa de desempenho, representada no eixo X. A altura da barra representa a quantidade de alunos formados que possuem desempenho dentro daquela faixa, de acordo com o eixo Y.

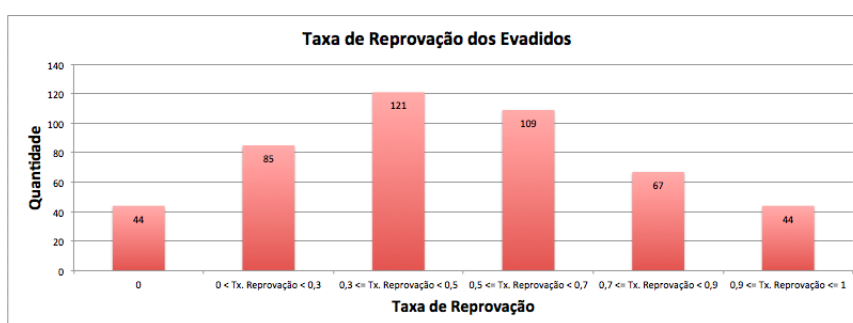


Figura 5.11: Taxa de reprovação média dos evadidos organizados em faixas de reprovação. Cada uma das barras corresponde a uma faixa de taxa de reprovação, representada no eixo X. A altura da barra representa a quantidade de alunos evadidos que possuem taxa de reprovação dentro daquela faixa, de acordo com o eixo Y.

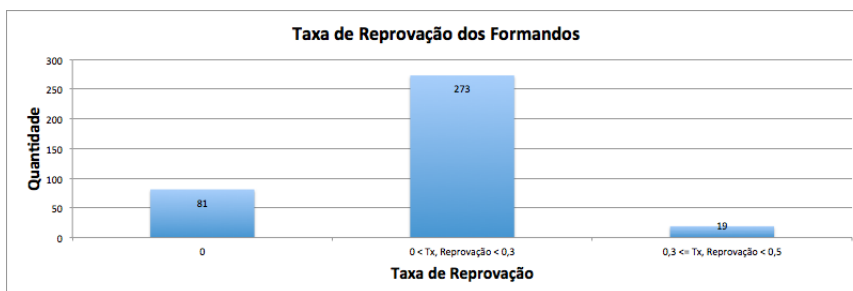


Figura 5.12: Taxa de reprovação média dos formados organizados em faixas de reprovação. Cada uma das barras corresponde a uma faixa de taxa de reprovação, representada no eixo X. A altura da barra representa a quantidade de alunos formados que possuem taxa de reprovação dentro daquela faixa, de acordo com o eixo Y.

5.5.12 Menções em disciplinas obrigatórias

As Tabelas C.26 e C.27 apresentam, respectivamente, os dados das menções nas disciplinas obrigatórias obtidas pelos alunos evadidos e formados. As Figuras 5.26, 5.27, 5.28, 5.29 e 5.30 apresentam os gráficos correspondentes, organizados de acordo com o semestre do fluxo que são recomendadas de serem cursadas. Foram selecionadas apenas disciplinas

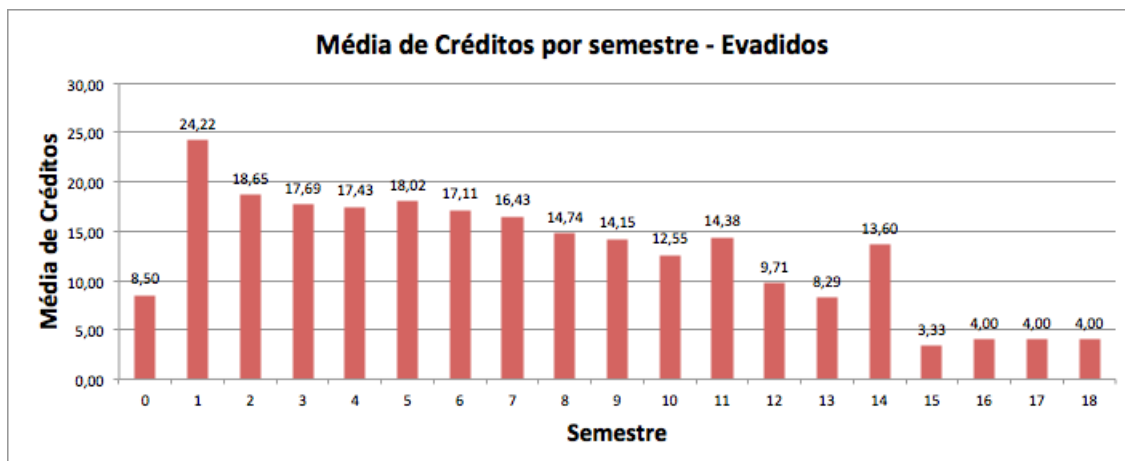


Figura 5.13: Média de créditos obrigatórios cursados por semestre dos evadidos. O semestre 0 representa um semestre cursado no verão, independente do período curricular que ele estava quando cursou. Cada uma das barras corresponde a um semestre, representada no eixo X. A altura da barra representa a média de créditos obrigatórios cursados por semestre, de acordo com o eixo Y.

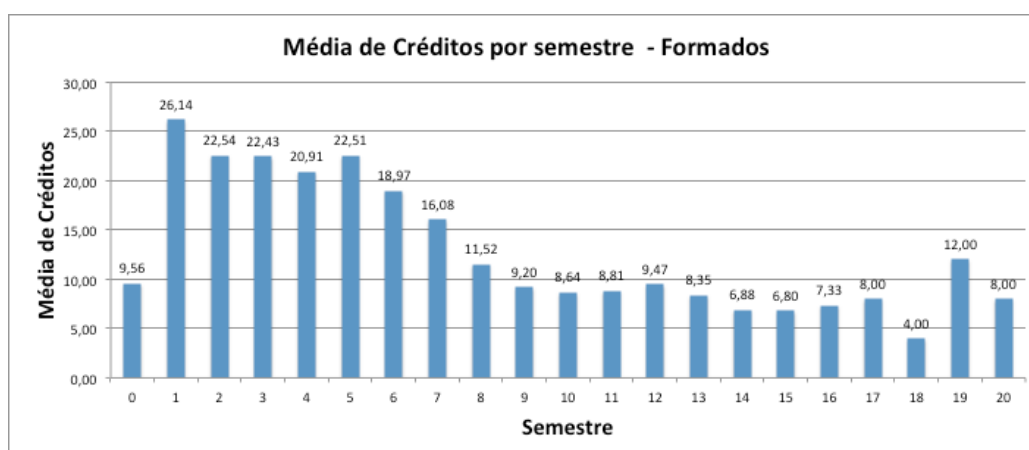


Figura 5.14: Média de créditos obrigatórios cursados por semestre dos formados. O semestre 0 representa um semestre cursado no verão, independente do período curricular que ele estava quando cursou. Cada uma das barras corresponde a um semestre, representada no eixo X. A altura da barra representa a média de créditos obrigatórios cursados por semestre, de acordo com o eixo Y.

com alta taxa de reprovação, ou seja, com valores maiores ou iguais a 25%, de acordo com os resultados obtidos na Subseção 5.5.10.

Podemos perceber que a quantidade de menções SR, TJ e TR é muito superior no caso dos alunos evadidos do que nos alunos formados. Mesmo em caso de aprovações, as menções obtidas pelos formados são mais altas, a maioria das menções dos formados é MS enquanto a dos evadidos é MM.

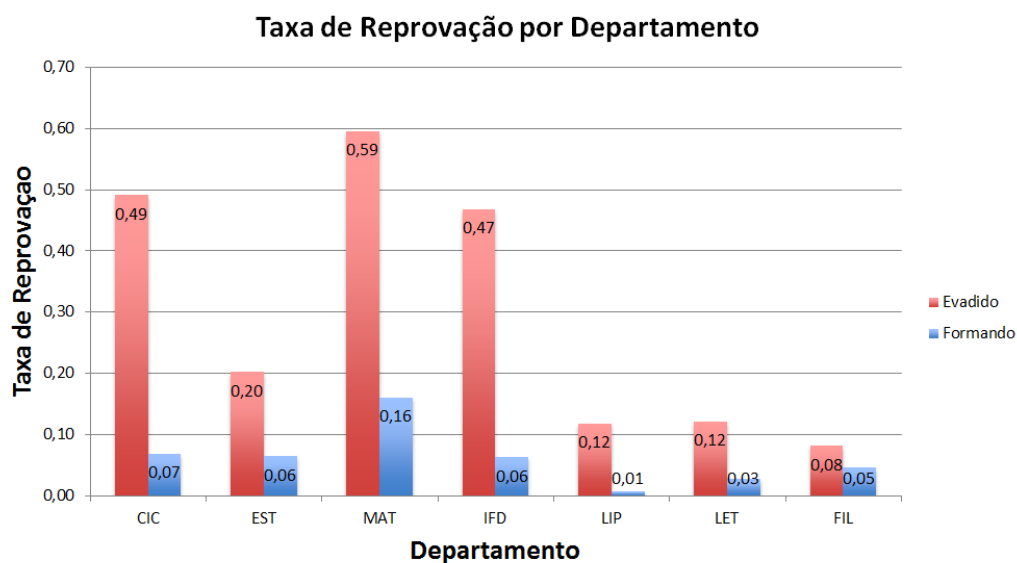


Figura 5.15: Índice de reprovação das matérias obrigatórias organizadas por departamento dos alunos evadidos e dos formados. Os departamentos estão representados no eixo X. A altura da barra representa a taxa de reprovação deles, independente da situação do estudante no curso, de acordo com o eixo Y.

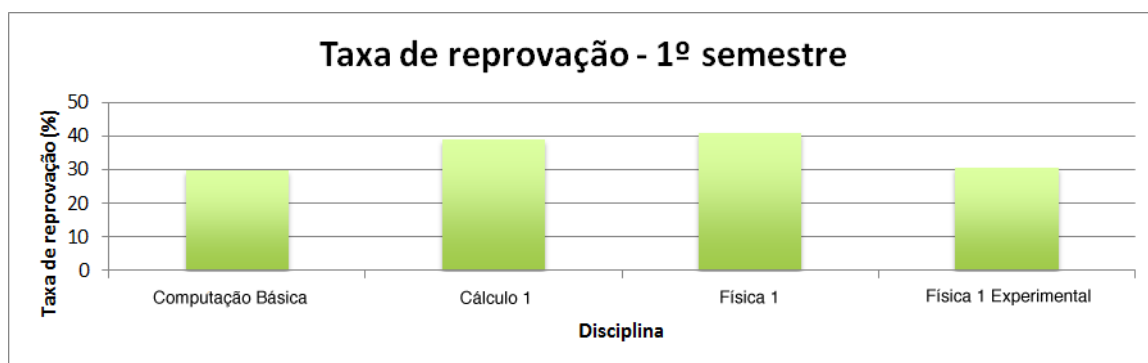


Figura 5.16: Reprovações nas disciplinas obrigatórias recomendadas de serem cursadas no 1º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10. Cada uma das barras corresponde a uma disciplina, representada no eixo X. A altura da barra representa a quantidade de reprovações que já ocorreram naquela disciplina, independente da situação do estudante no curso, de acordo com o eixo Y.

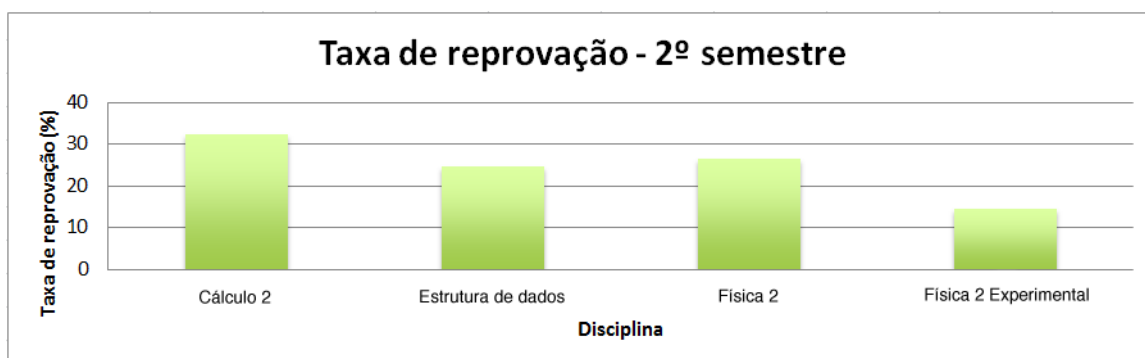


Figura 5.17: Reprovações nas disciplinas obrigatórias recomendadas de serem cursadas no 2º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10. Cada uma das barras corresponde a uma disciplina, representada no eixo X. A altura da barra representa a quantidade de reprovações que já ocorreram naquela disciplina, independente da situação do estudante no curso, de acordo com o eixo Y.



Figura 5.18: Reprovações nas disciplinas obrigatórias recomendadas de serem cursadas no 3º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10. Cada uma das barras corresponde a uma disciplina, representada no eixo X. A altura da barra representa a quantidade de reprovações que já ocorreram naquela disciplina, independente da situação do estudante no curso, de acordo com o eixo Y.

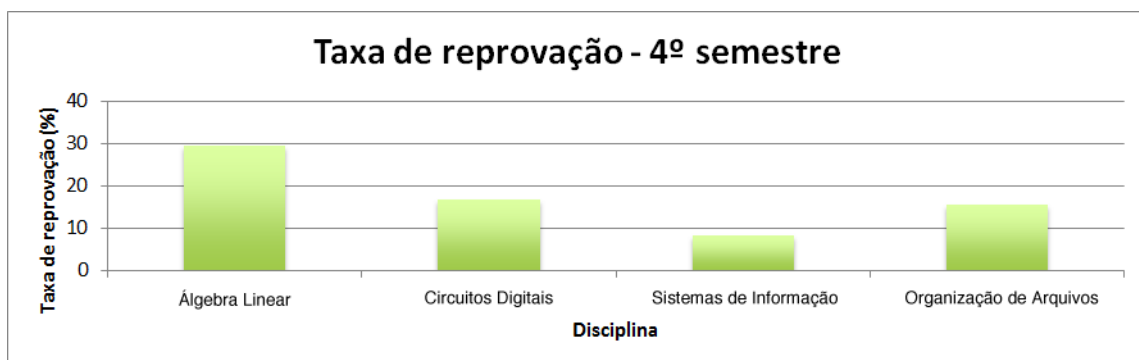


Figura 5.19: Reprovações nas disciplinas obrigatórias recomendadas de serem cursadas no 4º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10. Cada uma das barras corresponde a uma disciplina, representada no eixo X. A altura da barra representa a quantidade de reprovações que já ocorreram naquela disciplina, independente da situação do estudante no curso, de acordo com o eixo Y.

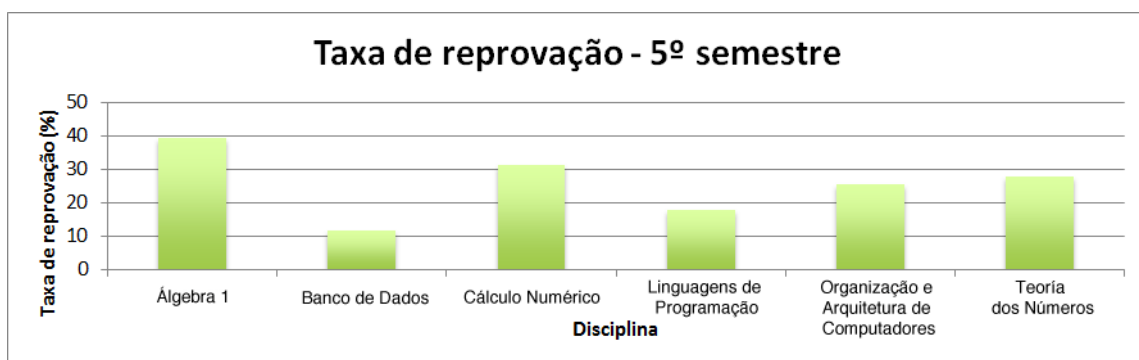


Figura 5.20: Reprovações nas disciplinas obrigatórias recomendadas de serem cursadas no 5º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10. Cada uma das barras corresponde a uma disciplina, representada no eixo X. A altura da barra representa a quantidade de reprovações que já ocorreram naquela disciplina, independente da situação do estudante no curso, de acordo com o eixo Y.

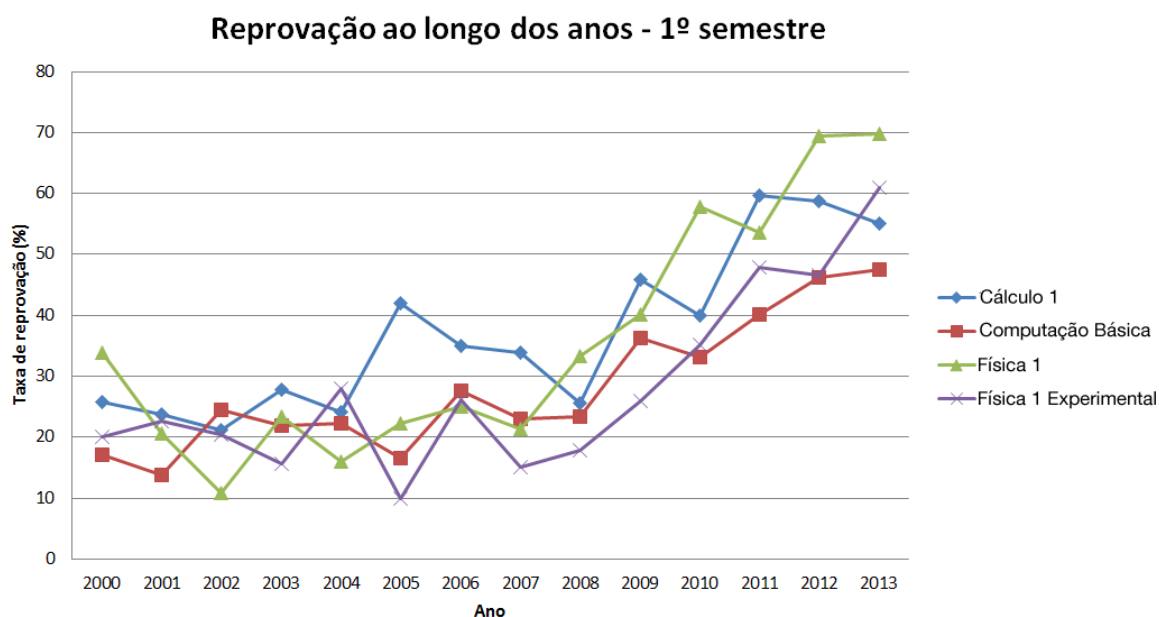


Figura 5.21: Evolução da taxa de reprovação ao longo dos anos nas disciplinas obrigatórias recomendadas de serem cursadas no 1º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10. Cada linha está relacionada a uma disciplina, de acordo com a legenda do lado direito do gráfico. Cada um dos pontos dessas linhas relaciona a taxa de reprovação em um determinado ano. A taxa de reprovação foi calculada como a proporção de reprovações das matrículas em uma determinada disciplina a cada ano.

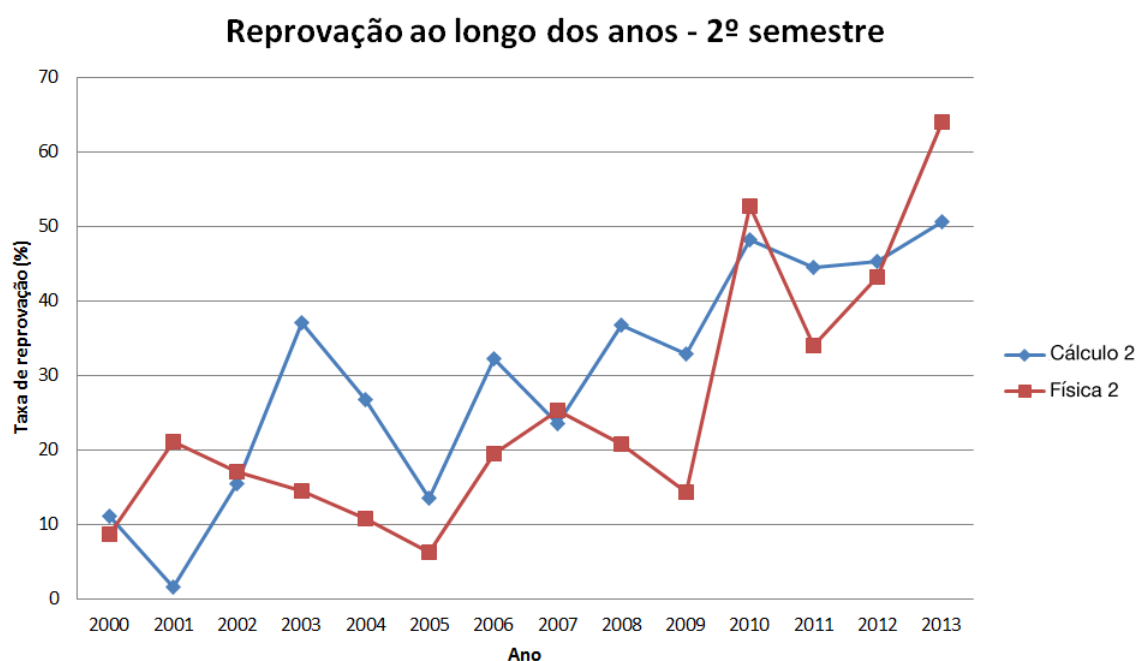


Figura 5.22: Evolução da taxa de reprovação ao longo dos anos nas disciplinas obrigatórias recomendadas de serem cursadas no 2º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10. Cada linha está relacionada a uma disciplina, de acordo com a legenda do lado direito do gráfico. Cada um dos pontos dessas linhas relaciona a taxa de reprovação em um determinado ano. A taxa de reprovação foi calculada como a proporção de reprovações das matrículas em uma determinada disciplina a cada ano.

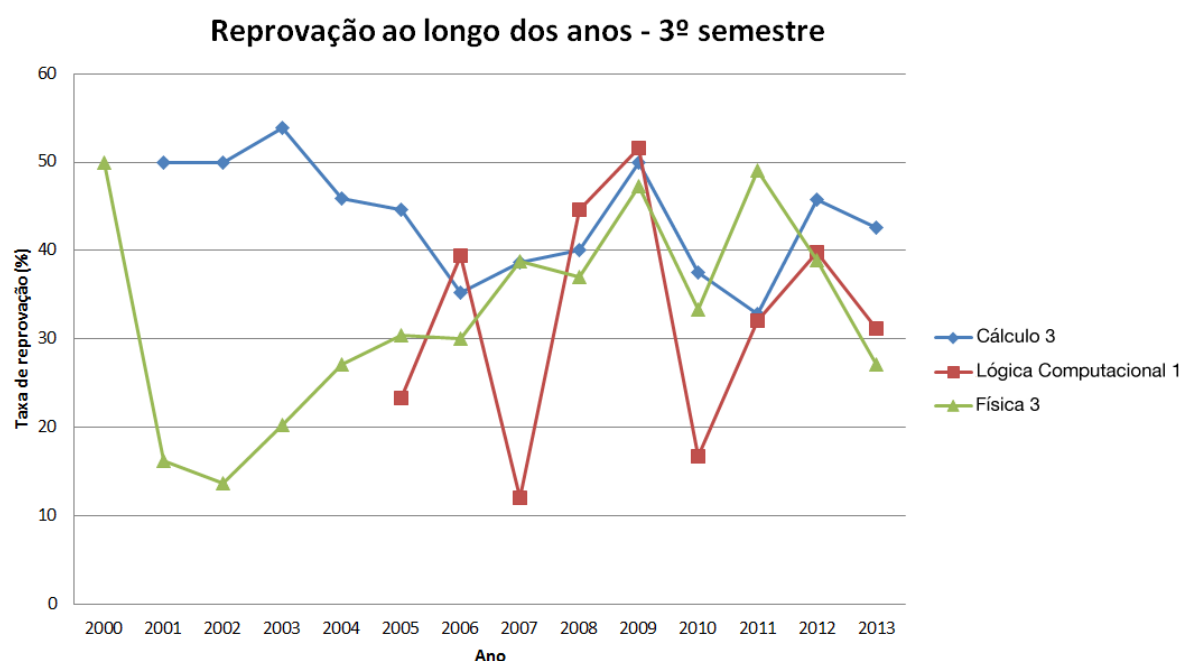


Figura 5.23: Evolução da taxa de reprovação ao longo dos anos nas disciplinas obrigatórias recomendadas de serem cursadas no 3º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10. Cada linha está relacionada a uma disciplina, de acordo com a legenda do lado direito do gráfico. Cada um dos pontos dessas linhas relaciona a taxa de reprovação em um determinado ano. A taxa de reprovação foi calculada como a proporção de reprovações das matrículas em uma determinada disciplina a cada ano.

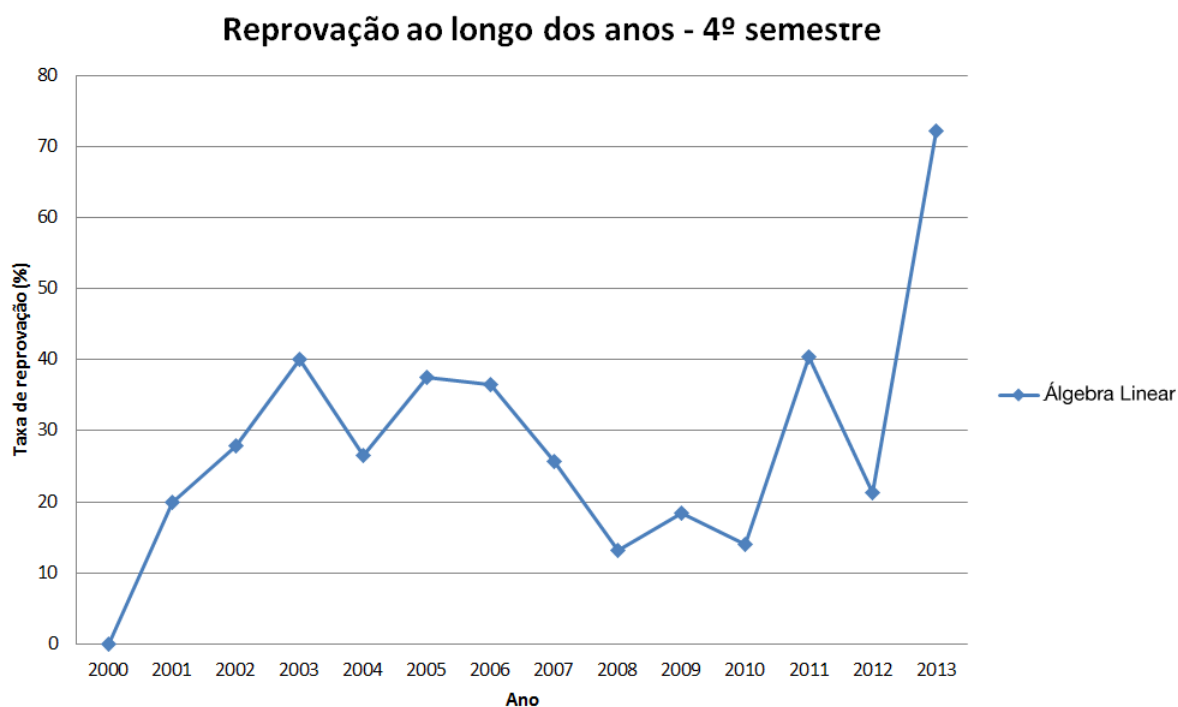


Figura 5.24: Evolução da taxa de reprovação ao longo dos anos nas disciplinas obrigatórias recomendadas de serem cursadas no 4º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10. Cada linha está relacionada a uma disciplina, de acordo com a legenda do lado direito do gráfico. Cada um dos pontos dessas linhas relaciona a taxa de reprovação em um determinado ano. A taxa de reprovação foi calculada como a proporção de reprovações das matrículas em uma determinada disciplina a cada ano.

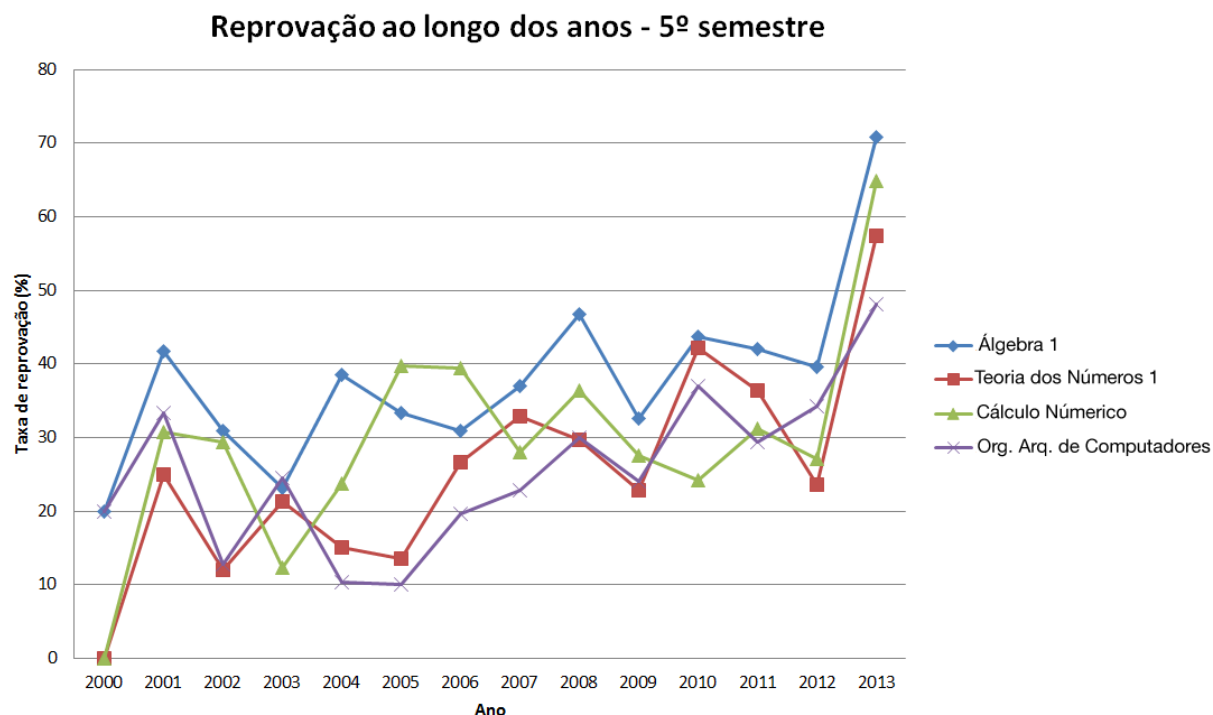


Figura 5.25: Evolução da taxa de reprovação ao longo dos anos nas disciplinas obrigatórias recomendadas de serem cursadas no 5º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10. Cada linha está relacionada a uma disciplina, de acordo com a legenda do lado direito do gráfico. Cada um dos pontos dessas linhas relaciona a taxa de reprovação em um determinado ano. A taxa de reprovação foi calculada como a proporção de reprovações das matrículas em uma determinada disciplina a cada ano.

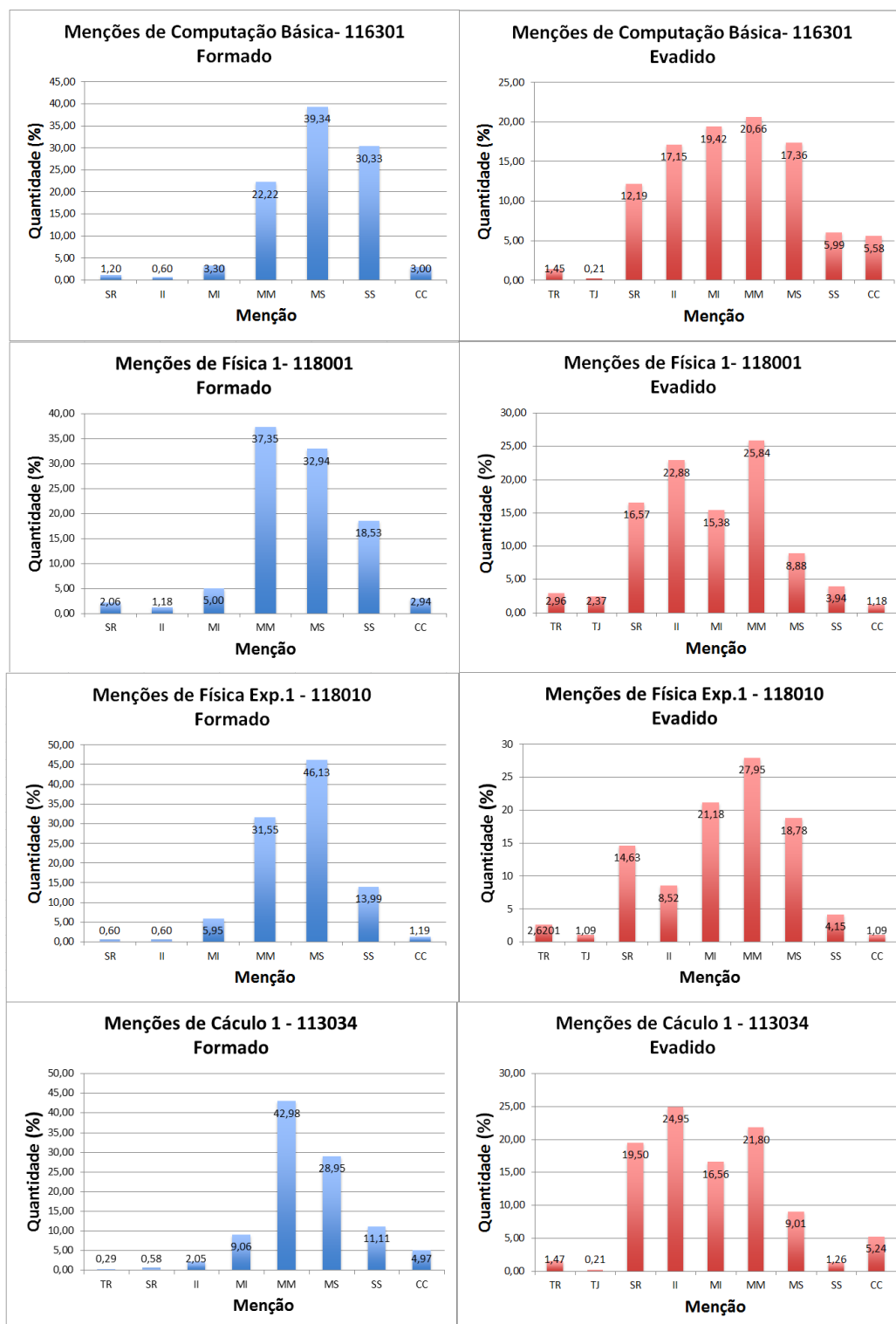


Figura 5.26: Menções nas disciplinas obrigatórias recomendadas de serem cursadas no 1º semestre do curso dos alunos formados e dos evadidos. Cada uma das barras corresponde a uma menção, representada no eixo X. A altura da barra representa a proporção de cada uma das menções obtidas naquela disciplina, de acordo com o eixo Y.

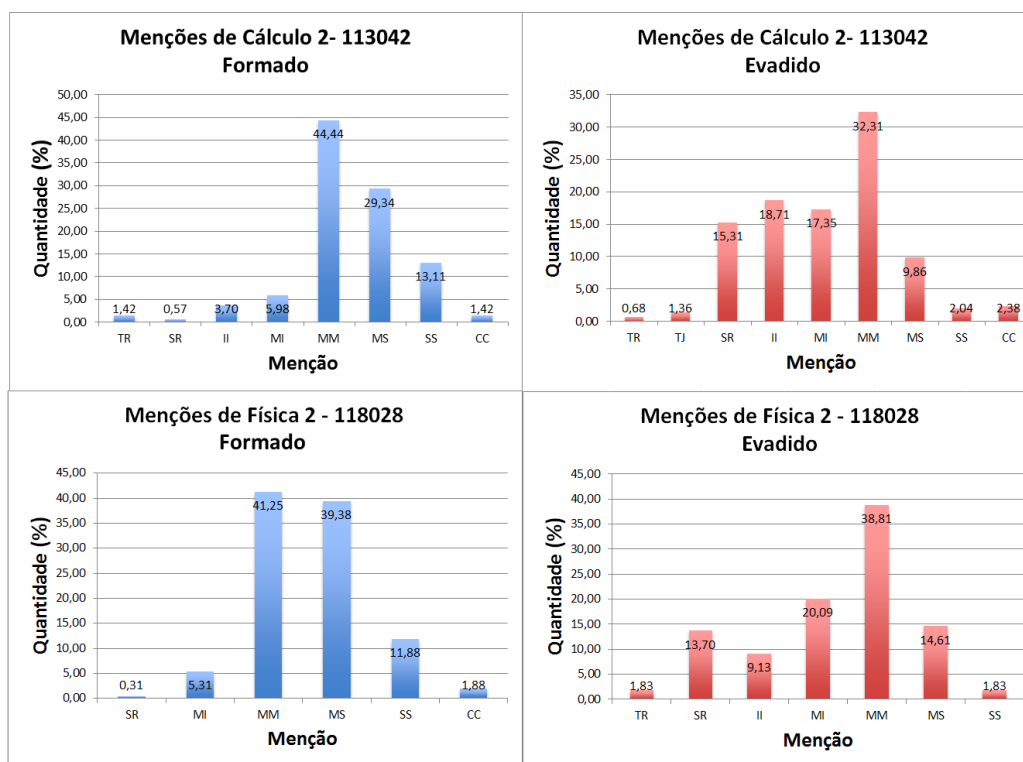


Figura 5.27: Menções nas disciplinas obrigatórias recomendadas de serem cursadas no 2º semestre do curso dos alunos formados e dos evadidos. Cada uma das barras corresponde a uma menção, representada no eixo X. A altura da barra representa a proporção de cada uma das menções obtidas naquela disciplina, de acordo com o eixo Y.

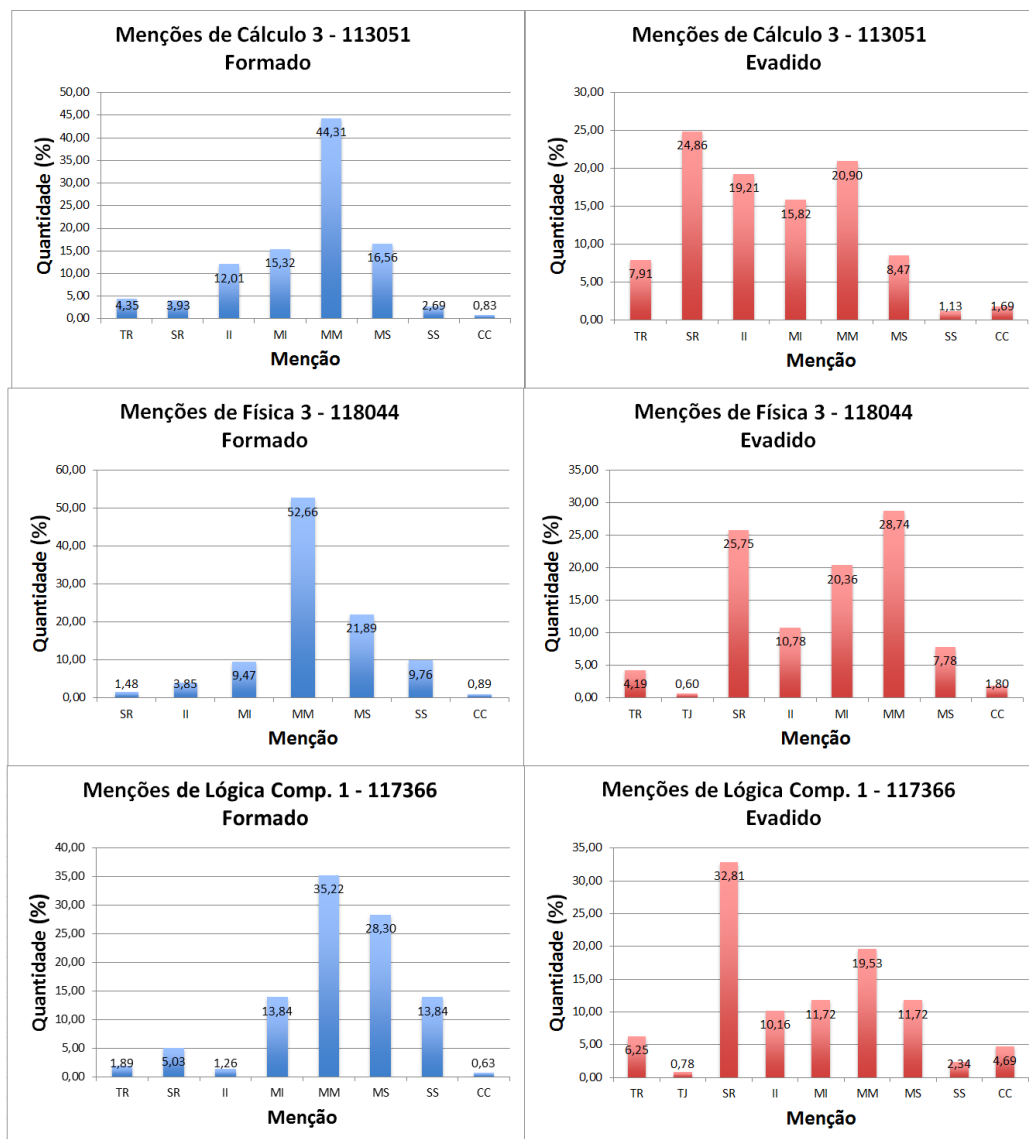


Figura 5.28: Menções nas disciplinas obrigatórias recomendadas de serem cursadas no 3º semestre do curso dos alunos formados e dos evadidos. Cada uma das barras corresponde a uma menção, representada no eixo X. A altura da barra representa a proporção de cada uma das menções obtidas naquela disciplina, de acordo com o eixo Y.

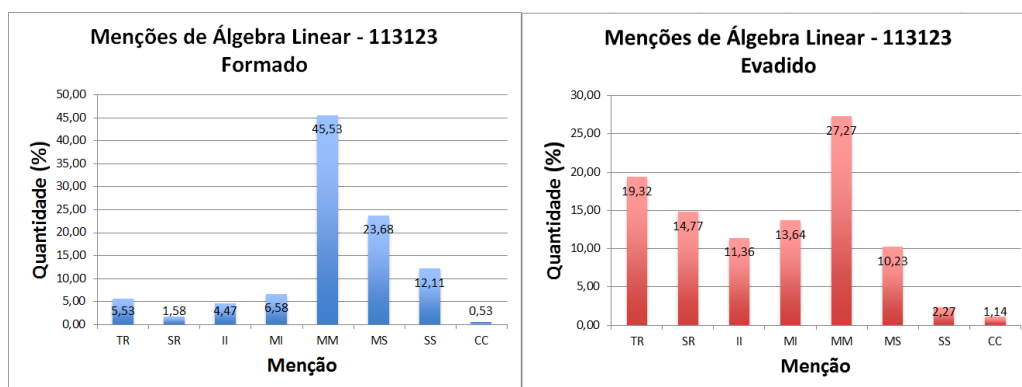


Figura 5.29: Menções nas disciplinas obrigatórias recomendadas de serem cursadas no 4º semestre do curso dos alunos formados e dos evadidos. Cada uma das barras corresponde a uma menção, representada no eixo X. A altura da barra representa a proporção de cada uma das menções obtidas naquela disciplina, de acordo com o eixo Y.

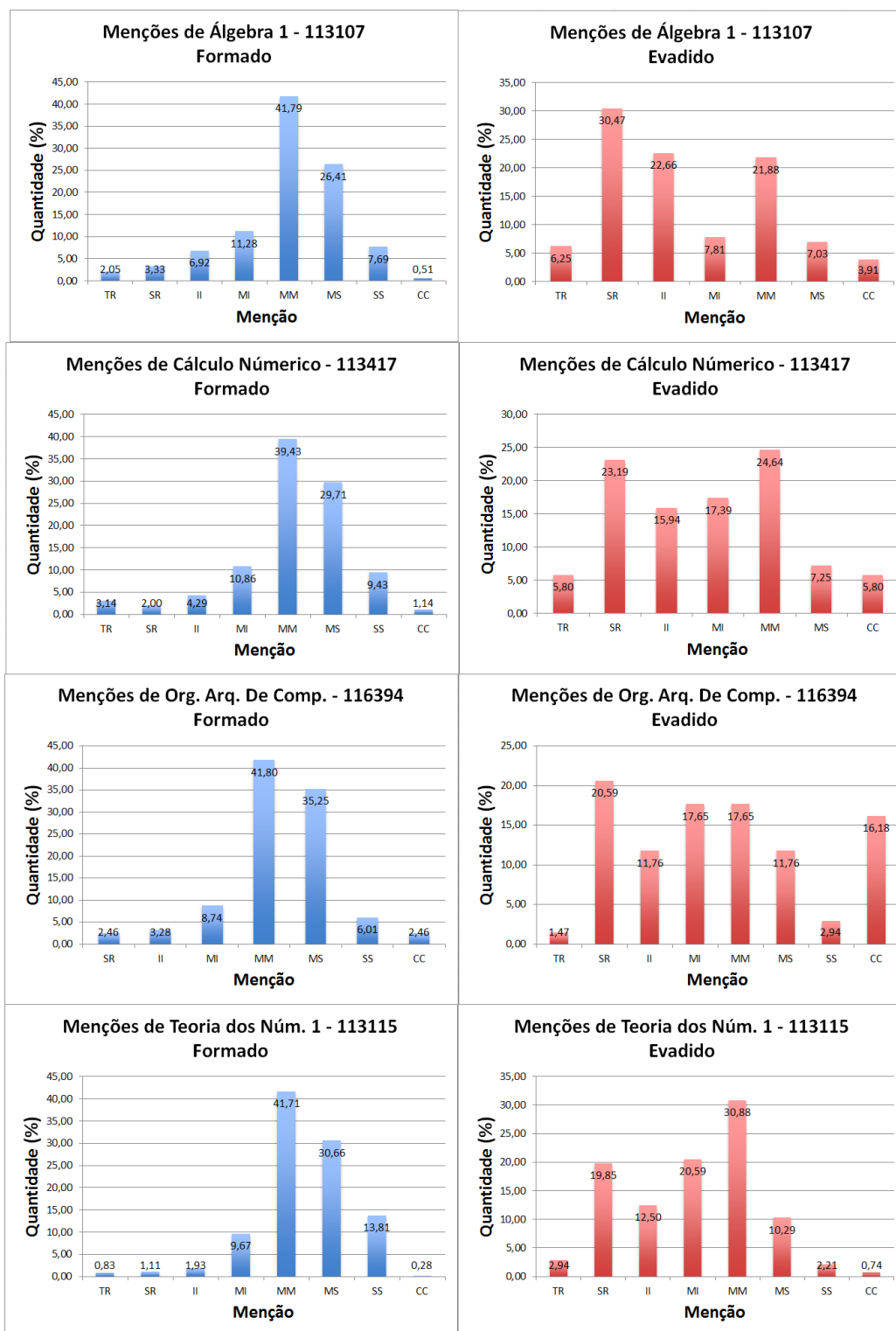


Figura 5.30: Menções nas disciplinas obrigatórias recomendadas de serem cursadas no 5º semestre do curso dos alunos formados e dos evadidos. Cada uma das barras corresponde a uma menção, representada no eixo X. A altura da barra representa a proporção de cada uma das menções obtidas naquela disciplina, de acordo com o eixo Y.

Capítulo 6

Perfil dos Alunos do BCC – UnB

Conforme citado no Capítulo 5, nesse capítulo é apresentado o processo da mineração de dados realizado para determinação do perfil dos alunos com risco de evasão.

Nesse capítulo, a Seção 6.1 apresenta a manipulação dos dados realizada para melhor adaptação para o processo de mineração de dados e as variáveis escolhidas; o processo da mineração é descrito na Seção 6.2; a Seção 6.3 apresenta os resultados das performances dos algoritmos utilizados para mineração. Com base nos resultados da mineração de dados, a Seção 6.4 apresenta como está a situação de alunos com perfil de evasão que estão cursando.

6.1 Transformação dos dados

Com base na análise estatística dos dados apresentada no Capítulo 5, foi possível determinar quais as variáveis possuem maior impacto na determinação do perfil de um aluno, que pode ser evadido ou formado.

As variáveis escolhidas, que serão utilizadas no processo de mineração de dados, são apresentadas na Tabela 6.1. O sexo foi considerado como uma variável importante, pois, apesar da proporção de homens que evade ser praticamente a mesma da de mulheres, os principais motivos de saída dos que evadem são distintos e, conseqüentemente, são perfis diferentes. Cada forma de ingresso na Universidade possui diferentes taxas de evasão, sendo que algumas chegam a 100% e outras a menos de 10%. A partir da análise estatística, foi concluído que os índices de reprovações nas matérias obrigatórias, em especial em alguns departamentos, são os principais diferenciadores dos dois tipos de perfis de alunos. Além disso, as menções obtidas nas disciplinas obrigatórias, tanto de aprovações quanto reprovações e trancamentos, estão fortemente ligadas com o perfil do aluno. Portanto, o IRAObr, calculado como uma relação entre a quantidade de cada menção obtida em disciplinas obrigatórias, sendo que quanto maior a menção, maior o peso nesse cálculo, também foi considerado uma variável importante.

Foram criadas duas novas tabelas no banco de dados com as informações apresentadas na Tabela 6.1. A primeira com dados dos alunos que já saíram do curso, seja por evasão ou formatura, e a segunda com os dados dos que estão cursando atualmente. Para poder ter a separação da situação do aluno (evadido, formado ou cursando), foi necessário adicionar nas duas tabelas o atributo ForSaidaOpcao. A consulta SQL utilizada para criação das tabelas no banco de dados pode ser verificada no Apêndice D.

Tabela 6.1: Variáveis escolhidas para serem utilizadas no processo de mineração de dados. A escolha das variáveis foi feita de acordo com os fatores que possuem maior impacto na definição do perfil, com base nos resultados obtidos na análise estatística dos dados, apresentada no Capítulo 5.

Variável	Descrição
ForIngressoOpcao	Forma de ingresso na Universidade.
AluSexo	Sexo do aluno.
IRAObr	Cálculo relacionado com as menções obtidas nas disciplinas obrigatórias que cursou, de forma que quanto maior a menção, maior o peso dela.
Cic	Taxa de reprovação nas disciplinas obrigatórias oferecidas pelo Departamento de Ciência da Computação.
Est	Taxa de reprovação nas disciplinas obrigatórias oferecidas pelo Departamento de Estatística.
Mat	Taxa de reprovação nas disciplinas obrigatórias oferecidas pelo Departamento de Matemática.
IFD	Taxa de reprovação nas disciplinas obrigatórias oferecidas pelo Instituto de Física.
Lip	Taxa de reprovação nas disciplinas obrigatórias oferecidas pelo Departamento de Lingüística, Português e Línguas Clássicas.
Let	Taxa de reprovação nas disciplinas obrigatórias oferecidas pelo Departamento de Línguas Estrangeiras e Tradução.
Fil	Taxa de reprovação nas disciplinas obrigatórias oferecidas pelo Departamento de Filosofia.

6.2 Mineração dos dados

A ferramenta Weka foi escolhida para ser utilizada no processo de mineração dos dados, pois, segundo Witten et al. [35] a ferramenta facilita o processo de mineração por já ter implementado os principais algoritmos de mineração, além de fornecer a performance do algoritmo e prever a classificação de novos dados.

O aprendizado selecionado foi o supervisionado. Essa escolha foi feita por conta do aprendizado supervisionado, segundo Donalek [13], permitir que o processo aprenda com base em exemplos de dados que já possuem os resultados esperados conhecidos. Nesse trabalho, já temos os resultados esperados dos alunos que não estão mais cursando. Esses dados serão utilizados para treinamento e teste.

Dentre as possíveis técnicas de mineração de dados com aprendizado supervisionado, a escolhida foi a classificação. Essa técnica, de acordo com Han e Kamber [19], permite

mapear os dados de entrada em classes já determinadas. O classificador irá aprender como determinar a classe de uma determinada entrada com base nos padrões que ele identificou na fase de treinamento. Assim como no trabalho realizado por Adhatrao et al. [1], a escolha da classificação foi feita considerando que o objetivo é classificar os alunos que estão cursando em duas classes já definidas: **evadido** ou **formado**.

Com base nos estudos de trabalhos anteriormente realizados, foram escolhidos os algoritmos de classificação mais utilizados. Assim, serão usados nesse projeto os algoritmos: *Naive Bayes*, *Support Vector Machine* (SVM), *k-Nearest Neighbor* (kNN), ID3 e C4.5.

O Weka permite diferentes formatos de dados de entrada, como planilhas, bancos de dados e geração de dados na própria ferramenta. No entanto, para determinados algoritmos e entradas, esses formatos geram erros. Para evitar esses erros, os dados de entrada devem ser colocados no formato ARFF (*Attribute-Relation File Format*). A partir de uma planilha no formato CSV (separado por vírgula), a conversão para o formato ARFF é realizada facilmente, de forma automática ou manual. A automática apresenta, em algumas situações, problema de incompatibilidade de dados das fases de treinamento, teste e classificação de novos dados. Esses problemas ocorrem por conta da existência de atributos com valores distintos dos apresentados na fase de treinamento. Portanto, a conversão manual é uma opção mais adequada para o processo. As etapas da conversão manual são:

1. salvar dados em uma planilha no formato CSV;
2. abrir arquivo CSV em um editor de texto;
3. inserir @relation <nome_relacao> no início do arquivo para definir o nome da relação entre os atributos;
4. inserir @attribute <nome_atributo> <{possiveis_valores}> para cada um dos atributos;
5. inserir @data para indicar que começará a listagem dos dados. Cada registro estará em uma linha e os dados são separados por vírgula, organizados na mesma ordem em que foram declarados os atributos.

Para realização do processo de mineração de dados foram gerados três arquivos no formato ARFF, com base nas duas novas tabelas criadas no banco de dados, apresentadas na Seção 6.1:

1. Parte dos dados (75%) dos alunos que já saíram do curso, formados ou evadidos. Esses dados serão usados para treinamento do classificador e teste para determinação da performance dele.
2. Parte dos dados (25%) dos alunos que já saíram do curso, formados ou evadidos. Esses dados serão usados para verificar se a acurácia obtida na fase de treinamento e de teste estão de acordo com a obtida utilizando novos dados de teste.
3. Dados dos alunos que estão cursando atualmente para verificar em qual perfil, evadido ou formado, eles se enquadram de acordo com o classificador utilizado.

```

1  @relation sairam_curso
2
3  @attribute ForIngressoOpcao {1,2,3,4,5,6,7,17,20,24}
4  @attribute ForSaidaOpcao {1,2,0}
5  @attribute AluSexo {M,F}
6  @attribute IRAObr {6,7,8,9,10,11}
7  @attribute Cic {10,11,12,13,14,15,16,17,18,19,20,21}
8  @attribute Est {10,11,12,13,14,15,16,17,18,19,20,21}
9  @attribute Mat {10,11,12,13,14,15,16,17,18,19,20,21}
10 @attribute IFD {10,11,12,13,14,15,16,17,18,19,20,21}
11 @attribute Lip {10,11,12,13,14,15,16,17,18,19,20,21}
12 @attribute Let {10,11,12,13,14,15,16,17,18,19,20,21}
13 @attribute Fil {10,11,12,13,14,15,16,17,18,19,20,21}
14
15 @data
16 1,1,M,10,11,10,12,10,10,10,10
17 1,1,M,11,10,10,10,10,10,10,10
18 1,1,M,11,11,10,10,10,10,10,16
19 1,2,M,8,15,16,17,17,10,10,16
20 1,2,M,8,15,10,21,10,10,21,21
21 1,1,M,11,10,10,10,10,10,10,10

```

Figura 6.1: Exemplo de arquivo no formato ARFF. O nome da relação entre os atributos é definido por @relation. Cada um dos atributos e seus possíveis valores são declarados por @attribute. O início da listagem dos dados é identificado por @data. Cada registro está em uma linha e os valores dos atributos são separados por vírgula e estão na mesma ordem da declaração dos atributos.

Os dados utilizados possuem valores numéricos e nominais. No entanto, alguns dos algoritmos escolhidos aceitam apenas dados nominais. Essa conversão pode ser feita diretamente no Weka, utilizando um filtro. As etapas para essa conversão são:

1. acessar Weka no modo *Explorer*;
2. abrir o arquivo no formato ARFF;
3. escolher o filtro;
4. seleccionar a opção *filters* → *unsupervised* → *attribute* → *NumericToNominal*;
5. aplicar o filtro;
6. salvar o arquivo.

A Figura 6.2 mostra as etapas da conversão de valores numéricos para nominais no Weka. Os números estão de acordo com os passos apresentados.

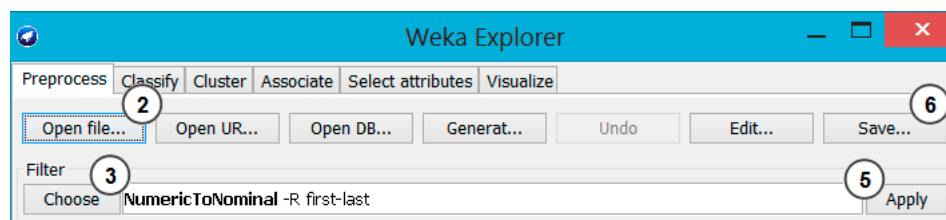


Figura 6.2: Conversão de valores numéricos para nominais utilizando o filtro do Weka. Os números estão de acordo com os passos apresentados.

Para criar um classificador utilizando a ferramenta Weka, é necessário seguir os seguintes passos:

1. abrir o Weka no modo *Explorer*;
2. selecionar o arquivo com os dados de teste e treinamento de acordo com o filtro necessário pelo algoritmo que será utilizado;
3. abrir a aba *Classify*;
4. selecionar o algoritmo que será utilizado. O nome da classe utilizado por cada um dos algoritmos é apresentado na Tabela 6.2;

Tabela 6.2: Nome dos algoritmos utilizados para classificação nesse trabalho e o nome da classe deles na ferramenta Weka.

Algoritmo	Nome da classe no Weka
Naive Bayes	bayes → NaiveBayes
k-Nearest Neighbors	functions → IBk
Support Vector Machine (SVM)	lazy → LibSVM
ID3	trees → Id3
C4.5	trees → j48

5. em *Test Options*, selecionar *Percentage Split* com valor 75%;
6. em *More options*, marcar a opção *Output prediction*;
7. selecionar o *Label*, que é o atributo utilizado para classificação dos dados. Escolher a opção *ForSaidaOpcao*;
8. clicar em *Start* para iniciar a classificação;
9. em *Result List*, clicar com o botão direito no classificador desejado;
10. selecionar a opção *Save model* e salvar o classificador.

6.3 Avaliação dos algoritmos

A performance dos classificadores é avaliada de acordo com critérios gerais do classificador e específicos de cada uma das classes. Os critérios gerais do classificador são aqueles referentes à acurácia e taxas de erro. Já os critérios específicos de cada classe se referem às taxas para determinar a probabilidade de uma instância ser classificada naquela classe. A Tabela 6.3 apresenta os critérios gerais. Os critérios específicos da classe de **formados** são apresentados na Tabela 6.4 e a de **evadidos** na Tabela 6.5.

Para verificar se a performance dos classificadores é mantida com dados distintos dos utilizados para treinamento e teste, é feita uma reavaliação do classificador utilizando novos dados de teste. Para isso, deve-se:

1. abrir o Weka no modo *Explorer*;

Tabela 6.3: Performance dos classificadores criados com base em critérios gerais do classificador.

Critério	Naive Bayes	SVM	kNN	ID3	C4.5
Correctly Classified Instances	90.0826%	83.4711%	88.4298%	90.9091%	85.124%
Kappa statistic	0.7893	0.6466	0.7526	0.8586	0.68
Mean absolute error	0.0766	0.1102	0.0897	0.0503	0.1444
Root mean squared error	0.2373	0.332	0.2544	0.2138	0.2755
Relative absolute error	23.4124%	33.6991%	27.443%	15.7943%	44.1568%
Root relative squared error	58.9584%	82.4648%	63.1984%	53.8709%	68.4366%
UnClassified Instances	0%	0%	0%	2.4793%	0%

Tabela 6.4: Performance dos classificadores criados com base em critérios específicos da classe **formados**.

Algoritmo	TP Rate	FP Rate	Precision	Recall	F-measure	ROC Area
Naive Bayes	0.986	0.22	0.864	0.986	0.921	0.961
SVM	0.944	0.32	0.807	0.944	0.87	0.812
kNN	0.986	0.26	0.843	0.986	0.909	0.888
ID3	0.957	0.104	0.931	0.957	0.944	0.915
C4.5	0.972	0.32	0.812	0.972	0.885	0.878

Tabela 6.5: Performance dos classificadores criados com base em critérios específicos da classe **evadidos**.

Algoritmo	TP Rate	FP Rate	Precision	Recall	F-measure	ROC Area
Naive Bayes	0.78	0.014	0.975	0.78	0.867	0.961
SVM	0.68	0.056	0.895	0.68	0.773	0.812
kNN	0.74	0.014	0.974	0.74	0.841	0.888
ID3	0.896	0.043	0.935	0.896	0.915	0.9
C4.5	0.68	0.028	0.944	0.68	0.791	0.878

- selecionar o arquivo com os dados de teste e treinamento de acordo com o filtro necessário pelo algoritmo que será utilizado;
- abrir a aba *Classify*;
- em *Result List*, clicar com o botão direito;
- selecionar a opção *Load Model*;
- selecionar o classificador desejado;
- em *Test options*, selecionar *Supplied test set*;

8. escolher o arquivo com os novos dados de teste de acordo com o filtro necessário pelo algoritmo que foi utilizado para criação do classificador;
9. em *Result List*, clicar com o botão direito no classificador desejado;
10. selecionar a opção *Re-evaluate model on current test set*.

A Tabela 6.6 apresenta a performance dos classificadores após o teste com novos dados com base nos critérios gerais. Os critérios específicos da classe de **formados** são apresentados na Tabela 6.7 e a de **evadidos** na Tabela 6.8.

Tabela 6.6: Performance dos classificadores criados com base em critérios gerais do classificador.

Critério	Naive Bayes	SVM	kNN	ID3	C4.5
Correctly Classified Instances	90.6977%	83.7209%	86.9767%	83.2558%	89.3023%
Kappa statistic	0.746	0.6023	0.6604	0.7063	0.7056
Mean absolute error	0.0802	0.1085	0.0941	0.0763	0.096
Root mean squared error	0.2458	0.3294	0.2744	0.2709	0.2416
Relative absolute error	—	29.9528%	—	—	—
Root relative squared error	—	74.0123%	—	—	—
UnClassified Instances	0%	0%	0%	6.5116%	0%

Tabela 6.7: Performance dos classificadores após o teste com novos dados com base em critérios específicos da classe **formados**.

Algoritmo	TP Rate	FP Rate	Precision	Recall	F-measure	ROC Area
Naive Bayes	1	0.115	0.672	1	0.804	0.977
SVM	1	0.201	0.539	1	0.701	0.899
kNN	0.976	0.155	0.597	0.976	0.741	0.924
ID3	0.95	0.124	0.655	0.95	0.776	0.898
C4.5	0.951	0.121	0.65	0.951	0.772	0.878

Nesse experimento, o principal fator de avaliação de um algoritmo é sua acurácia, ou seja, qual a porcentagem de instâncias classificadas corretamente. Assim, no primeiro teste realizado, o *Naive Bayes* e o ID3 apresentaram os melhores resultados. O desempenho desses algoritmos também pode ser verificado por meio do critério *Kappa statistic*, que possui os maiores valores dentre os testes realizados, e das menores taxas de erro. Apesar desses critérios para o classificador com algoritmo ID3 serem melhores, uma desvantagem dele foi a dificuldade que ele apresentou de classificar cerca de 2.48% das instâncias.

Ao rodar novos testes sobre o classificador, apenas a acurácia do *Naive Bayes* se manteve, a do ID3 reduziu em, aproximadamente, 7%. Por isso, o classificador construído com base no algoritmo *Naive Bayes* obteve a melhor performance nesse experimento.

Tabela 6.8: Performance dos classificadores após o teste com novos dados com base em critérios específicos da classe **evadidos**.

Algoritmo	TP Rate	FP Rate	Precision	Recall	F-measure	ROC Area
Naive Bayes	0.885	0	1	0.885	0.939	0.977
SVM	0.799	0	1	0.799	0.888	0.899
kNN	0.845	0.024	0.993	0.845	0.913	0.924
ID3	0.876	0.05	0.986	0.876	0.928	0.865
C4.5	0.879	0.049	0.987	0.879	0.93	0.954

Analisando os resultados do *Naive Bayes* para cada uma das classes, é possível concluir que esse classificador acerta mais o perfil de evasão dos alunos.

6.4 Alunos com perfil de evasão

Como o classificador criado com base no algoritmo *Naive Bayes* apresentou a melhor performance, ele será utilizado para identificar o perfil dos alunos que estão cursando atualmente.

A classificação é feita utilizando um arquivo de teste que possui dados apenas dos alunos que estão cursando. Os seguintes passos são necessários para essa classificação.

1. abrir o Weka no modo *Explorer*;
2. selecionar o arquivo com os dados de teste e treinamento de acordo com o filtro necessário pelo algoritmo que será utilizado;
3. abrir a aba *Classify*;
4. em *Result List*, clicar com o botão direito;
5. selecionar a opção *Load Model*;
6. selecionar o classificador construído com o algoritmo *Naive Bayes*;
7. em *Test options*, selecionar *Supplied test set*;
8. escolher o arquivo com os dados dos alunos cursando de acordo com o filtro necessário pelo algoritmo que foi utilizado para criação do classificador;
9. em *Result List*, clicar com o botão direito no classificador construído com o algoritmo *Naive Bayes*;
10. selecionar a opção *Re-evaluate model on current test set*.

Após isso, é gerada a *Confusion Matrix*, em que é possível visualizar como foram classificados os dados. A matriz mostra como os alunos cursando (representado por $c = 0$) foram classificados, podendo ser em formados ($a = 1$) ou evadidos ($b = 2$). Como na fase de treinamento foram utilizados apenas dados de formados ou evadidos e o objetivo desse trabalho é classificar os alunos nesses dois perfis, a matriz, que aparenta estar errada já que nenhum dado foi classificado corretamente, na verdade mostra uma predição de como esses dados seriam classificados. A Figura 6.3 apresenta essa matriz gerada.

```

a   b   c   <-- classified as
0   0   0   |   a = 1
0   0   0   |   b = 2
109 154  0   |   c = 0

```

Figura 6.3: *Confusion matrix* gerada pelo classificador construído com o algoritmo *Naive Bayes*. Essa matriz mostra quantos alunos, dos que estão cursando, são classificados como **formandos** e quantos como **evadidos**. O número total de instâncias é 263.

Com base nos dados da matriz apresentada na Figura 6.3, foi gerado o gráfico da Figura 6.4, que mostra a proporção de alunos cursando que foram classificados como formandos ou evadidos.

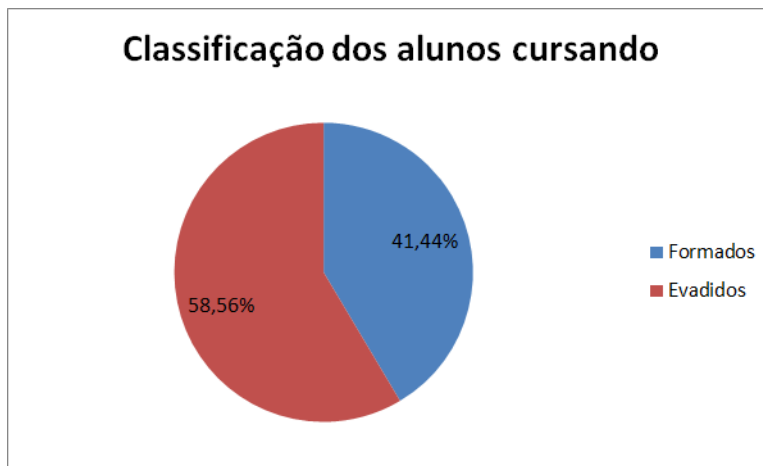


Figura 6.4: Proporção de alunos cursando que foram classificados como formandos ou evadidos pelo classificador construído com o algoritmo *Naive Bayes*.

Dos alunos cursando, 58.56% foram classificados com perfil de evasão. Esse número, em comparação com a taxa de evasão dos que já saíram que é de 55.76%, se mostra realista uma vez que o desempenho nas disciplinas foi considerado o principal determinante para o perfil de evasão e, conforme mostrado na análise estatística dos dados apresentada no Capítulo 5, o índice de reprovações nas disciplinas obrigatórias estão aumentando cada vez mais nos últimos quatro anos, o que representa grande parte dos dados de alunos cursando atualmente.

Capítulo 7

Conclusões

Os resultados experimentais obtidos nesse trabalho validam a hipótese de que é possível utilizar técnicas de mineração de dados para determinar o perfil de um aluno em risco de evasão no BCC – UnB.

A análise estatística dos dados foi fundamental para identificação das principais variáveis relacionadas com o perfil de evasão ou de formatura de um aluno. Foi verificado que a taxa de reprovação nas disciplinas obrigatórias do curso está de certa forma relacionada com a decisão por evasão do aluno, especialmente quando a reprovação ocorre por trancamento ou por falta de rendimento.

Construindo classificadores com diferentes algoritmos, foi verificado que o *Naive Bayes* apresentou o melhor resultado para os atributos escolhidos nesse trabalho. Esse classificador determinou que 58.56% dos alunos cursando atualmente irão sair do curso sem se formar. Atualmente a taxa de evasão é de 55.76%, mostrando que o problema está se agravando no curso.

7.1 Principais contribuições

As principais contribuições desse trabalho foram:

1. Análise estatística de aspectos ligados à evasão do BCC - UnB.
2. Geração de banco de dados contendo dados filtrados dos alunos de 2000 a 2013, que pode ser utilizado em trabalhos futuros.
3. Compreensão do perfil e das características do aluno que podem influenciar na sua evasão.
4. Criação de um classificador que automaticamente consegue prever a saída por evasão ou por formatura de um determinado aluno.

Com base nessas contribuições, é possível que professores e coordenadores identifiquem os alunos que possuem perfil de evasão e, com isso, sejam capazes de fazer uma intervenção para melhorar seu ensino com intuito de incentivar que o aluno se forme no curso. Os dados e estatísticas geradas também permitem a geração de conhecimento para melhoria da abordagem das disciplinas oferecidas pelo Departamento de Ciência da Computação.

7.2 Trabalhos futuros

Esse trabalho representa uma base no estudo do perfil de evasão dos alunos do Departamento de Ciência da Computação. Portanto, diferentes novos trabalhos podem ser realizados para aprofundamento dos resultados obtidos. Possíveis trabalhos são:

1. Analisar a situação dos alunos que não entraram diretamente no BCC – UnB.
2. Construir um novo classificador que faça uso de dados das disciplinas optativas, além dos atributos já utilizados nesse trabalho.
3. Utilizar técnicas de clusterização para identificar outros possíveis perfis de alunos.
4. Aplicar métodos de regressão nos atributos com a finalidade de gerar indicadores de relações entre esses e o atributo pergunta.
5. Desenvolver um sistema com interface amigável para gerenciamento dos dados pelo coordenador do curso, chefe do departamento ou funcionários de outros setores na Universidade.
6. Repetir o experimento para o curso de Computação - Licenciatura, que também é oferecido pelo Departamento de Ciência da Computação.
7. Repetir o experimento futuramente para verificar mudanças de perfil por conta do novo PPP do curso.

Referências

- [1] K. Adhatrao, A. Gaykar, A. Dhawan, R. Jha, and V. Honrao. Predicting students' performance using iD3 and c4.5 classification algorithms. *International Journal of Data Mining Knowledge Management Process (IJDKP)*, 3(5), 2013. 22, 71
- [2] E. Alpaydin. *Introduction to Machine Learning*. The MIT Press, 2nd edition, 2010. 19, 22
- [3] F. Araque, C. Roldán, and A. Salguero. Factors influencing university drop out rates. *Computers & Education*, 53(3):563–574, 2009. 26, 33, 34, 35, 36
- [4] C. Baggi and D. Lopes. Evasão e avaliação institucional no ensino superior: uma discussão bibliográfica. *Avaliação: Revista da Avaliação da Educação Superior (Campinas)*, 16(2):355–374, 2011. 6
- [5] J. Bertolin. Indicadores em nível de sistema para avaliar o desenvolvimento e a qualidade da educação superior brasileira. *Avaliação: Revista da Avaliação da Educação Superior (Campinas)*, 12:309 – 331, 06 2007. 5, 6
- [6] A. de V. C. Campello and L. N. Lins. Metodologia de análise e tratamento da evasão e retenção em cursos de graduação instituições federais de ensino superior. *Anais XXVIII Encontro Nacional de Engenharia de Produção, Rio de Janeiro*, 2008. xiii, 29, 30, 31, 33, 34, 35, 36
- [7] A. Casartelli, A. C. Benso, M. Morosini, and R. M. Gessinger. Um estudo sobre os motivos relacionados à evasão estudantil na PUCRS. *Segunda Conferencia sobre el Abandono en la Educación Superior (II CLABES)*, pages 221 – 227. 7
- [8] G. Couto and M. Dantas. Utilizando mineração de dados na resolução de problemas de gênero na UnB. 2014. Monografia do curso de Computação – Licenciatura – Universidade de Brasília, Brasília. 28
- [9] Universidade de Brasilia – UnB. Desligamentos, 2014. Disponível em: <http://www.saa.unb.br/index.php/desligamento> Acessado em: 16/06/2014. 37
- [10] Universidade de Brasília – UnB. Currículo do Bacharelado em Ciência da Computação. Disponível em: <https://matriculaweb.unb.br/matriculaweb/graduacao/curriculo.aspx?cod=1856> Acessado em: 07/06/2014. xiii, xiv, 9, 10, 11, 83
- [11] Universidade de Brasília – UnB. Curso de Ciência da Computação. Disponível em: http://www.unb.br/aluno_de_graduacao/cursos/ciencia_da_computacao Acessado em: 07/06/2014. 8

- [12] INEP – Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Sinopses estatísticas da educação básica, 2012. Disponível em: <http://portal.inep.gov.br/basica-censo-escolar-sinopse-sinopse> Acessado em: 09/04/2014. **xiii, 4, 5**
- [13] C. Donalek. Supervised and unsupervised learning, 2011. Disponível em: http://www.astro.caltech.edu/~george/aybi199/ek_Classif.pdf Acessado em: 01/06/2014. **19, 20, 22, 70**
- [14] P. dos Santos. Evasão na educação superior: uma análise a partir de publicações na ANPED e CAPES (2000 a 2012). *Tercera Conferencia sobre el Abandono en la Educación Superior (III CLABES)*, pages 413 – 421. **7**
- [15] P. dos Santos and L. Giraffa. Evasão na educação superior: um estudo sobre o censo da educação superior no brasil. *Tercera Conferencia sobre el Abandono en la Educación Superior (III CLABES)*, pages 49 – 58. **6**
- [16] U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, editors. *Advances in Knowledge Discovery and Data Mining*. American Association for Artificial Intelligence, Menlo Park, CA, USA, 1996. **15, 17**
- [17] J. B. Ferreira, M. de L. Machado-Taylor, and A. Magalhães. A importância e a satisfação no ensino superior: a perspectiva dos estudantes. In *X Congresso da Sociedade Portuguesa de Ciências da Educação*. Instituto Politécnico de Bragança, 2009. **1**
- [18] V. Guidotti and P. Verdum. Fatores que influenciam a evasão e a permanência dos alunos de um curso de pedagogia na modalidade EAD. *Tercera Conferencia sobre el Abandono en la Educación Superior (III CLABES)*, pages 249 – 258. **6**
- [19] J. Han and M. Kamber. *Data mining : concepts and techniques*. Kaufmann, San Francisco, 2005. **vii, 14, 15, 16, 17, 18, 19, 20, 21, 22, 70**
- [20] A. de O. Lima and B. A. C. Costa. Evasão na Universidade de Brasília: um estudo sobre os cursos formadores de docentes para matérias básicas do vestibular. 2012. Monografia (Bacharelado em Estatística) – Universidade de Brasília, Brasília. **6, 27, 28, 33, 34, 35, 36**
- [21] M. B. Lobo. Panorama da evasão no ensino superior brasileiro: aspectos gerais das causas e soluções. Disponível em: http://www.institutolobo.org.br/imagens/pdf/artigos/art_087.pdf Acessado em: 09/04/2014. **6, 37**
- [22] H. Martins Galvão, H. L. Corrêa, and J. L. Alves. Modelo de avaliação de desempenho global para instituição de ensino superior. *Revista de Administração da Universidade Federal de Santa Maria [On-line]*, 4:425–441, Novembro 2011. **4**
- [23] M. Morosini, A. Casartelli, A. C. da Silva, B. dos Santos, R. Schmitt, and R. M. Gessinger. A evasão na Educação Superior no Brasil: uma análise da produção de conhecimento nos periódicos qualis entre 2000-201. *Primera Conferencia Latinoamericana sobre el Abandono en la Educación Superior (I CLABES)*, pages 65 – 74. **7, 27, 28**

- [24] Weka – The University of Waikato. Weka 3: Data Mining Software in Java, 2013. Disponível em: <http://www.cs.waikato.ac.nz/ml/weka/index.html> Acessado em: 01/06/2014. 23
- [25] S. Perdomo, G. Labra, and I. Nogueira. Evasão na Escola Superior de Saúde da Universidade do Estado do Amazonas. *Segunda Conferencia sobre el Abandono en la Educación Superior (II CLABES)*, pages 119 – 127. 7
- [26] A. K. Pujari. *Data mining techniques*. Universities Press, 2001. 19
- [27] R. Roiger and M. W. Geatz. *Data Mining: A Tutorial Based Primer*. Addison Wesley, 2003. 22
- [28] E. O. Santos and V. C. Valverde. A evasão no curso de Química da UnB: o que mudou após 1997. 2010. Monografia (Especialização em Desenvolvimento Gerencial) – Universidade de Brasília, Brasília. 27, 33, 34, 35, 36
- [29] F. Sarker, H. Davis, and T. Tiropanis. A review of higher education challenges and data infrastructure responses. In *International Conference for Education Research and Innovation (ICERI2010)*. International Conference for Education Research and Innovation (ICERI2010), November 2010. Event Dates: 15th -17th November 2010. 1
- [30] F. Sarker, T. Tiropanis, and H. Davis. Exploring student predictive model that relies on institutional databases and open data instead of traditional questionnaires. In *Proceedings of the 22Nd International Conference on World Wide Web Companion, WWW '13 Companion*, pages 413–418, Republic and Canton of Geneva, Switzerland, 2013. International World Wide Web Conferences Steering Committee. 8
- [31] H. R. B. da Silva and P. J. L. Adeodato. A data mining approach for preventing undergraduate students retention. In *Neural Networks (IJCNN), The 2012 International Joint Conference on*, pages 1–8. IEEE, 2012. 33, 34, 35, 36
- [32] R. Silva Filho, P. R. Motejunas, O Hipólito, and M.B. Lobo. A evasão no ensino superior brasileiro. *Cadernos de Pesquisa*, 37(132):641–659, 2007. 6
- [33] I. M. de Souza et al. Causas da evasão nos cursos de graduação da Universidade Federal de Santa Catarina. 1999. Dissertação (Mestrado em Administração – Universidade Federal de Santa Catarina, Florianópolis. 27, 28, 33, 34, 35, 36
- [34] R. F. Vitelli, C. S. Rocha, and R. Fritsch. Estudo sobre evasão nos cursos de graduação de uma instituição de ensino superior privada: Aplicação de regressão logística. xiii, 29, 30, 33, 34, 35, 36
- [35] I. H. Witten, E. Frank, and M. A. Hell. *Data Mining: Practical Machine Learning Tools and Technique*. Morgan Kaufmann, 3 edition, 2011. 14, 23, 70
- [36] X. Wu, V. Kumar, J. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z. Zhou, M. Steinbach, D. J. Hand, and D. Steinberg. Top 10 algorithms in data mining. *Knowl. Inf. Syst.*, 14(1):1–37, December 2007. 21

Anexo I

Ementas e pré-requisitos das disciplinas obrigatórias

A Tabela I.1 apresenta a lista das ementas das disciplinas obrigatórias do BCC – UnB e seus pré-requisitos.

Tabela I.1: Ementas e pré-requisitos das disciplinas obrigatórias do BCC – UnB [10].

Código da disciplina	Pré-requisitos	Ementa
113107	-	<ul style="list-style-type: none">- Noções de lógica e teoria dos conjuntos;- Relações e funções;- Noções de reticulados e álgebras de Boole;- Axiomas de Peano;- Princípios de indução- Os números inteiros; congruências;- Estruturas algébricas: semi-grupos, grupos, anéis, domínios de integridade e corpos- Polinômios sobre domínios de integridade em uma variável
113123	113034	<ul style="list-style-type: none">- O espaço \mathbb{R}^n e o cálculo vetorial em \mathbb{R}^2 e \mathbb{R}^3- Corpos e espaços vetoriais- A teoria dos determinantes- Aplicações lineares e matrizes- Formas canônicas
116378	116327	<ul style="list-style-type: none">- Níveis de abstração de informações e dados- Modelo relacional; modelo de rede, modelo hierárquico- Segurança, integridade e concorrência- Estudos de caso- Novas aplicações para banco de dados

Continua na página seguinte

Tabela I.1 – *Continuação da tabela da página anterior*

Código da disciplina	Pré-requisitos	Ementa
113034	-	<ul style="list-style-type: none"> - Funções de uma variável real - Limite e continuidade - Derivada - Integral - Aplicações de integral
113042	113034	<ul style="list-style-type: none"> - Sequências e séries numéricas - Séries de potências - Fórmula de Taylor - Equações diferenciais ordinárias de 1ª ordem - Equações diferenciais ordinárias lineares - A Transformada de Laplace - Sistemas lineares de equações diferenciais ordinárias de 1ª ordem
113051	113042	<ul style="list-style-type: none"> - Vetores no plano e no espaço - Funções de várias variáveis - Fórmula de Taylor, pontos de extremos locais e absolutos, pontos críticos, Multiplicador de Lagrange - Transformações diferenciáveis - Integrais múltiplas - Integrais de linha - Integrais de superfícies, Teorema da Divergência e Teorema de Stokes
113417	113042	<ul style="list-style-type: none"> - Zero de funções reais - Sistemas de equações lineares e inversão de matrizes - Ajustes de curvas e interpolação - Integração numérica - Soluções numéricas de equações diferenciais ordinárias
116351	118044 E 118052 E 167037	<ul style="list-style-type: none"> - Representação da informação - Álgebra Booleana - Estrutura de portas lógicas - Operações com dados - Flip-flops e registradores - Circuitos sequenciais - Conversão D/A e A/D

Continua na página seguinte

Tabela I.1 – *Continuação da tabela da página anterior*

Código da disciplina	Pré-requisitos	Ementa
116301	-	<ul style="list-style-type: none"> - Histórico do computador - Computadores e resolução de problemas - Estruturas de decisão - Vetores e matrizes - Cadeias de caracteres - Subalgoritmos: funções e procedimentos - Estilo de programação - Particularidades da linguagem Pascal
116441	116343	<ul style="list-style-type: none"> - Software e engenharia de software - Planejamento de software - Fundamentos e métodos de análise - Fundamentos de projeto de software - Garantia da qualidade do software - Automação do processo de desenvolvimento de software
116319	116301 OU 117234	<ul style="list-style-type: none"> - Manipulação de cadeias - Estruturas de dados lineares: vetores, matrizes, pilhas e listas encadeadas - Estruturas de dados não-lineares: árvores, matrizes esparsas e grafos - Classificação e pesquisa em memória
118001	-	<ul style="list-style-type: none"> - Conceitos e operações básicas relativos à cinemática e à dinâmica dos movimentos de translação e rotação - Leis de Newton - Energia e potência - Equilíbrio de corpos rígidos - Colisões
118010	-	<ul style="list-style-type: none"> - Medidas e erros - Análise gráfica - Atrito - Colisão - Conservação do momento linear - Estudo dos movimentos - Rotação - Conservação de energia - Equilíbrio de corpos rígidos

Continua na página seguinte

Tabela I.1 – *Continuação da tabela da página anterior*

Código da disciplina	Pré-requisitos	Ementa
118028	118001 E 118010 E 113034	<ul style="list-style-type: none"> - Dinâmica da rotação - Conservação do momento angular - Oscilações - Gravitação - Estática dos fluidos - Dinâmica dos fluidos - Ondas e meios elásticos - Ondas sonoras - Temperatura - Calor e 1ª Lei da Termodinâmica - Teoria cinética dos gases - Entropia e 2ª Lei da Termodinâmica
118036	118001 E 118010 E 113034	<ul style="list-style-type: none"> - Giroscópio - Movimento periódico - Hidrostática - Ondas sonoras - Dilatação linear - Calor específico dos sólidos - Condução de calor - Comportamento dos gases
118044	118028 E 118036 E 113042 OU 118206 E 118214 E 113042	<ul style="list-style-type: none"> - Leis de Coulomb - Campo elétrico - Lei de Gauss - Potencial, capacitância, propriedade dos dielétricos - Corrente, resistência e Fem - Circuitos e instrumentos de corrente contínua - Campo magnético - Forças magnéticas sobre condutores de correntes - Campo magnético produzido por correntes - Força eletromotriz induzida - Correntes alternadas - Equações de Maxwell

Continua na página seguinte

Tabela I.1 – *Continuação da tabela da página anterior*

Código da disciplina	Pré-requisitos	Ementa
118052		- Fundamentos teóricos
		- Uso de medidores elétricos e fontes
		- Resistência elétrica
		- Ohmica e não ohmica
		- Fonte de tensão
		- Fem e resistência interna
	118028 E	- Ponte de Wheatstone
	118036 E	- Medidas de resistências, superfícies equipotenciais e linhas de campo
118052	113042 OU	- Deflexão eletromagnética de elétrons
	118206 E	- Capacitores sob correntes CC
	118214 E	- Força magnética sobre segmentos retilíneos de correntes
	113042	- As bobinas de Helmholtz e a razão e/m para o elétron
		- Circuitos com correntes alternadas
145971	-	- Ressonância no circuito RLC série
		- Transformadores e circuitos retificadores
145971	-	- Introdução e prática de estratégias de compreensão escrita que favoreçam uma leitura mais eficiente e independente de textos variados
		- Desenvolvimento da percepção dos princípios lógicos envolvidos no processo da leitura
116416	116319 OU 113956	- Fundamentos da teoria geral de sistemas
		- Teoria da informação: conceito de informação, conceitos de dados, representação de dados e de conhecimento
		- Sistemas de informação: fases e etapas
		- Documentação
		- Prototipação
		- Modelagem conceitual: abstração, modelo entidade-relacionamento, análise funcional, administração de dados
		- Estudo de caso

Continua na página seguinte

Tabela I.1 – *Continuação da tabela da página anterior*

Código da disciplina	Pré-requisitos	Ementa
113930	-	<ul style="list-style-type: none"> - Árvores - Conexão - Grafos eulerianos e hamiltonianos - Teoria transversão - Planaridade - Dualidade - Matroides - Teoremas de Hall, König e Kuratowski - Aplicações - Problema de otimização em grafos - Otimização - Algoritmos
140481	-	<ul style="list-style-type: none"> - Leitura ativa - Leitura analítica - Leitura crítica - Planejamento e produção de resumos, resenhas, críticas e textos dissertativos argumentativos
116343	116319	<ul style="list-style-type: none"> - Introdução - Conceitos: variáveis, expressões, escopo, comandos, tipagem, procedimentos, tipos de dados, abstração, controle, etc - Paradigmas e linguagem de programação: imperativas, lógicas e orientadas por objetos
116394	116351 OU 169251 OU 169102 OU 167983	<ul style="list-style-type: none"> - Desempenho de processadores - Arquitetura do conjunto de instruções e linguagem de máquina - Aritmética computacional - Organização e projeto de processadores (uniciclo, multiclo e pipeline) - Hierarquia de memória - Sistemas de entrada e saída - Introdução ao multiprocessamento
116327	115045 E 116319	<ul style="list-style-type: none"> - Tecnologias de dispositivos de armazenamento secundário - Parâmetros de hardware - Organizações básicas de arquivos - Pilha, sequencial, sequencial-indexado, direto e invertido - Organizações híbridas de arquivos: avaliação de sistemas de arquivos - Compressão de dados - Classificação em memória secundária

Continua na página seguinte

Tabela I.1 – *Continuação da tabela da página anterior*

Código da disciplina	Pré-requisitos	Ementa
115045	113034 OU 113018	<ul style="list-style-type: none"> - Análise de observações - Modelo matemático - Exp. Aleatória - Espaço amostral - Variáveis aleatórias - Distribuições e suas características - Covariância e correlação - Distribuição conjunta - Principais modelos - Discretos e contínuos - Estatística descritiva - Ajustamento de funções reais - Correlação e regressão - Noções de amostragem - Testes de hipóteses
113956	116301 OU 113913	<ul style="list-style-type: none"> - Especificação e definição de programas - Métodos de programação - Documentação - Testes sistemáticos - Manutenção de programas - Estudo de caso
116467	116432	<ul style="list-style-type: none"> - Conceitos básicos e históricos - Gerência de processos e programação concorrente - Gerência de dispositivos - Gerência de memória - Sistema de arquivos - Estudo de sistemas operacionais existentes
116432	116394 E 116319	<ul style="list-style-type: none"> - Programação em linguagem de máquina - Montadores, ligadores e carregadores - Programação em linguagem C
113115	-	<ul style="list-style-type: none"> - Indução matemática e Princípio da Boa Ordenação - Divisibilidade de inteiros - Equações Diofantinas - Teorema Fundamental da Aritmética e aplicações - Congruências - Tópicos adicionais

Continua na página seguinte

Tabela I.1 – *Continuação da tabela da página anterior*

Código da disciplina	Pré-requisitos	Ementa
116459		- Tipos de tradutores
		- Especificação de linguagens de programação
	113948 E	- Análise léxica
	116343 E	- Análise sintática
	116432 OU	- Verificação de tipos de análise de escopo
	116882 E	- Ambiente de execução
	116343 E	- Organização da memória
	116432	- Acesso a variáveis, passagem de parâmetros, tabelas de símbolos
		- Geração de código intermediário e código final

Apêndice A

Consulta SQL para tratamento dos dados

A seguinte consulta SQL foi utilizada para a etapa de tratamento dos dados.

```
SELECT * FROM 'historico' WHERE 'Disciplina' = 113107 OR 'Disciplina' = 113123 OR 'Disciplina' = 116378 OR 'Disciplina' = 113034 OR 'Disciplina' = 113042 OR 'Disciplina' = 113051 OR 'Disciplina' = 113417 OR 'Disciplina' = 116351 OR 'Disciplina' = 116301 OR 'Disciplina' = 116441 OR 'Disciplina' = 116319 OR 'Disciplina' = 118001 OR 'Disciplina' = 118010 OR 'Disciplina' = 118028 OR 'Disciplina' = 118036 OR 'Disciplina' = 118044 OR 'Disciplina' = 118052 OR 'Disciplina' = 145971 OR 'Disciplina' = 116416 OR 'Disciplina' = 113930 OR 'Disciplina' = 140481 OR 'Disciplina' = 116343 OR 'Disciplina' = 116394 OR 'Disciplina' = 116327 OR 'Disciplina' = 115045 OR 'Disciplina' = 113956 OR 'Disciplina' = 116467 OR 'Disciplina' = 116432 OR 'Disciplina' = 113115 OR 'Disciplina' = 116459 OR 'Disciplina' = 113948 OR 'Disciplina' = 116360 OR 'Disciplina' = 116882 OR 'Disciplina' = 116475 OR 'Disciplina' = 116912 OR 'Disciplina' = 116921 OR 'Disciplina' = 137481 OR 'Disciplina' = 117366 OR 'Disciplina' = 113859 OR 'Disciplina' = 117536 OR 'Disciplina' = 116424 OR 'Disciplina' = 204315

SELECT * FROM alunos INNER JOIN historico ON alunos.MatricAluno = historico.MatricAlunos WHERE alunos.FormaSaida = 9

UPDATE historico INNER JOIN optativas on historico.Disciplina = optativas.CodDisciplina SET historico.TipoDisciplina = 2

UPDATE historico INNER JOIN alunos on historico.MatricAluno = alunos.MatricAluno SET historico.TipoDisciplina = 2

UPDATE 'alunos' SET 'SemestreCursoSaida' = ('PerSaidaOpcao' - 'AnoIngresso')*2 + 1 WHERE 'SemestreIngresso' = 1 AND 'SemestreSaidaOpcao' = 1

SELECT count(MatricAluno) AS num_alunos, SemestreCursoSaida FROM alunos WHERE FormaSaida = 2 GROUP BY SemestreCursoSaida
```

```

SELECT count(MatricAluno) AS num_alunos, SemestreCursoSaida ,
    (count(MatricAluno)/462)*100 FROM alunos WHERE FormaSaida = 2 GROUP BY
    SemestreCursoSaida

SELECT AluSexo, count(MatricAluno) AS num_alunos FROM alunos WHERE
    FormaSaida = 2 GROUP BY 'AluSexo'

SELECT count(MatricAluno) AS num_alunos, IdadeSaida ,
    (count(MatricAluno)/462)*100 FROM alunos WHERE FormaSaida = 2 GROUP BY
    IdadeSaida

SELECT 'AluEscola', count(MatricAluno) AS num_alunos,
    (count(MatricAluno)/462)*100 FROM alunos WHERE FormaSaida = 2 GROUP BY
    AluEscola

SELECT 'FormaIngresso', count(MatricAluno) AS num_alunos,
    (count(MatricAluno)/462)*100 FROM alunos WHERE FormaSaida = 2 GROUP BY
    FormaIngresso

SELECT SUM historico.creditos HAVING historico.Menciao != ... / SUM
    (historico.creditos) FROM alunos INNER JOIN historico ON
    alunos.MatricAluno = historico.MatricAlunos WHERE alunos.FormaSaida = 2

SELECT (
    Select SUM(historico.creditos) FROM alunos INNER JOIN historico ON
        alunos.MatricAluno = historico.MatricAluno Where alunos.FormaSaida = 2
and historico.Aprovacao = 1 )
    / SUM(historico.creditos) FROM alunos INNER JOIN historico ON
        alunos.MatricAluno = historico.MatricAluno WHERE alunos.FormaSaida = 2

SELECT alunos.MatricAluno, SUM(historico.creditos), (Select
    SUM(historico.creditos) FROM alunos INNER JOIN historico ON
        alunos.MatricAluno = historico.MatricAluno Where alunos.FormaSaida = 2
and historico.Aprovacao = 1 )), ( Select SUM(historico.creditos) as
    aprovado FROM alunos INNER JOIN historico ON alunos.MatricAluno =
        historico.MatricAluno Where alunos.FormaSaida = 2
and historico.Aprovacao = 1 ) / SUM(historico.creditos) FROM alunos INNER
JOIN historico ON alunos.MatricAluno = historico.MatricAluno WHERE
    alunos.FormaSaida = 2 GROUP BY alunos.MatricAluno

SELECT SemestreCursoSaida, MotivoFormaSaidaOpcao, count(MatricAluno) AS
    num_alunos, count(MatricAluno)/4.74 as porcentagem, (select
    MAX(num_alunos) from alunos WHERE ForSaidaOpcao = 2 Group by
    SemestreCursoSaida)
FROM alunos WHERE ForSaidaOpcao = 2 GROUP BY SemestreCursoSaida,
    MotivoFormaSaidaOpcao

Select aprovadas.MatricAluno, creditosApro/creditotsTot as desempenho from (
SELECT alunos.MatricAluno, SUM(historico.creditos) as creditotsTot FROM
    alunos INNER JOIN historico ON alunos.MatricAluno =
        historico.MatricAluno WHERE alunos.FormaSaida = 2 and
        alunos.MatricAluno= 200035939 Group by alunos.MatricAluno) as total
inner join (

```

```

SELECT alunos.MatricAluno, SUM(historico.creditos) as creditosApro FROM
    alunos INNER JOIN historico ON alunos.MatricAluno =
    historico.MatricAluno WHERE historico.Aprovacao = 1 and
    alunos.FormaSaida = 2 and alunos.MatricAluno= 200035939 Group by
    alunos.MatricAluno) as aprovadas
on aprovadas.MatricAluno=total.MatricAluno
Group By aprovadas.MatricAluno

update alunos inner join
    (Select aprovadas.MatricAluno, creditosApro/creditosTot as desempenho from
        (
SELECT alunos.MatricAluno, SUM(historico.creditos) as creditosTot FROM
            alunos INNER JOIN historico ON alunos.MatricAluno =
            historico.MatricAluno Group by alunos.MatricAluno) as total
inner join (
SELECT alunos.MatricAluno, SUM(historico.creditos) as creditosApro FROM
            alunos INNER JOIN historico ON alunos.MatricAluno =
            historico.MatricAluno WHERE historico.Aprovacao = 1 Group by
            alunos.MatricAluno) as aprovadas
on aprovadas.MatricAluno=total.MatricAluno
Group By aprovadas.MatricAluno) as DesempenhoAlunos
on DesempenhoAlunos.MatricAluno=alunos.MatricAluno
set alunos.Desempenho=DesempenhoAlunos.desempenho

/*Separando os alunos que entraram diretamente em cic dos que n o
entraram:*/
Select Count(MatricAluno) From alunos Where AnoIngresso !=
    PerIngressoOpcao Or SemestreIngresso != SemestreIngressoOpcao

/*Semestre cursado*/
Update alunos
INNER JOIN historico ON alunos.MatricAluno= historico.MatricAluno
Set historico.SemetreCursado = (2*(historico.Ano-alunos.AnoIngresso)+ 1)
Where alunos.GrupoAnalise = 0 And historico.Semestre = 1 And
    alunos.SemestreIngresso = 1

Update alunos
INNER JOIN historico ON alunos.MatricAluno= historico.MatricAluno
Set historico.SemetreCursado = (2*(historico.Ano-alunos.AnoIngresso)+ 2)
Where alunos.GrupoAnalise = 0 And historico.Semestre = 2 And
    alunos.SemestreIngresso = 1

Update alunos
INNER JOIN historico ON alunos.MatricAluno= historico.MatricAluno
Set historico.SemetreCursado = (2*(historico.Ano-alunos.AnoIngresso)+ 1)
Where alunos.GrupoAnalise = 0 And historico.Semestre = 2 And
    alunos.SemestreIngresso = 2

update alunos inner join
    (Select aprovadas.MatricAluno, aprovadas.creditosApro/total.creditosTot as
        desempenho from

        (SELECT alunos.MatricAluno, SUM(historico.creditos) as creditosTot FROM
            alunos INNER JOIN historico ON alunos.MatricAluno =
            historico.MatricAluno JOIN obrigatorias ON obrigatorias.CodDisciplina =
            historico.Disciplina Group by alunos.MatricAluno) as total

```



```

inner join
(SELECT alunos.MatricAluno, SUM(historico.creditos) as creditosApro FROM
    alunos INNER JOIN historico ON alunos.MatricAluno =
    historico.MatricAluno JOIN obrigatorias ON obrigatorias.CodDisciplina =
    historico.Disciplina where historico.Aprovacao = 1 Group by
    alunos.MatricAluno) as aprovadas
on aprovadas.MatricAluno=total.MatricAluno
Group By aprovadas.MatricAluno) as DesempenhoAlunos
on DesempenhoAlunos.MatricAluno=alunos.MatricAluno
set alunos.Desempenho=DesempenhoAlunos.desempenho

/*TaxaReprovacao*/
update alunos inner join
(Select reprovadas.MatricAluno, reprovadas.creditosRepro/total.creditosTot
    as taxa from
    (SELECT alunos.MatricAluno, SUM(historico.creditos) as creditosTot FROM
        alunos INNER JOIN historico ON alunos.MatricAluno =
        historico.MatricAluno JOIN obrigatorias ON obrigatorias.CodDisciplina =
        historico.Disciplina Group by alunos.MatricAluno) as total
    inner join
    (SELECT alunos.MatricAluno, SUM(historico.creditos) as creditosRepro FROM
        alunos INNER JOIN historico ON alunos.MatricAluno =
        historico.MatricAluno JOIN obrigatorias ON obrigatorias.CodDisciplina =
        historico.Disciplina where historico.Aprovacao = 0 Group by
        alunos.MatricAluno) as reprovadas
    on reprovadas.MatricAluno=total.MatricAluno
    Group By reprovadas.MatricAluno) as ReprovacaoAlunos
on ReprovacaoAlunos.MatricAluno=alunos.MatricAluno
set alunos.TaxaReprovacao=ReprovacaoAlunos.taxa

```

Apêndice B

Consulta SQL para análise estatística dos dados

A seguinte consulta SQL foi utilizada para obter os resultados utilizados na análise estatística dos dados.

```
/*SEMESTRE*/  
SELECT SemestreCursoSaida , count(MatricAluno) FROM 'alunos' WHERE  
    'ForSaidaOpcao' = 1 GROUP BY SemestreCursoSaida  
  
SELECT  
    SemestreCursoSaida ,  
        MotivoFormaSaidaOpcao ,  
        count(MatricAluno) AS num_alunos ,  
        count(MatricAluno)/4.70 as porcentagem  
FROM alunos  
WHERE ForSaidaOpcao = 2  
GROUP BY SemestreCursoSaida , MotivoFormaSaidaOpcao  
Order by SemestreCursoSaida , num_alunos DESC  
  
/*IDADE*/  
SELECT  
    IdadeSaida ,  
    MotivoFormaSaidaOpcao ,  
    count(MatricAluno) AS num_alunos ,  
    count(MatricAluno)/4.70 as porcentagem  
FROM alunos  
WHERE ForSaidaOpcao = 2  
GROUP BY IdadeSaida , MotivoFormaSaidaOpcao  
Order by IdadeSaida , num_alunos DESC  
  
SELECT  
    IdadeSaida ,  
    MotivoFormaSaidaOpcao ,  
    count(MatricAluno) AS num_alunos ,  
    count(MatricAluno)*100/ total.AluTotal as porcentagem  
FROM alunos , (select count(MatricAluno) as AluTotal from alunos WHERE  
    ForSaidaOpcao = 2) as total  
WHERE ForSaidaOpcao = 2  
GROUP BY IdadeSaida , MotivoFormaSaidaOpcao  
Order by IdadeSaida , num_alunos DESC
```

```

/*SEXO*/
SELECT
    alunos.AluSexo,    count(MatricAluno) AS NumEvadidos,    TotAlunos
from alunos
inner Join
    ( SELECT
        AluSexo,    count(MatricAluno) AS TotAlunos
    FROM alunos
    GROUP BY 'AluSexo '
    ) as total
    On total.AluSexo = alunos.AluSexo
WHERE ForSaidaOpcao = 2
GROUP BY 'AluSexo '

SELECT
    alunos.AluSexo, MotivoFormaSaidaOpcao, count(MatricAluno) AS NumEvadidos
from alunos
WHERE ForSaidaOpcao = 2
GROUP BY alunos.'AluSexo ', alunos.MotivoFormaSaidaOpcao
Order by alunos.'AluSexo 'DESC

/*TIPO DE ESCOLA*/
SELECT
    alunos,AluEscola,    count(MatricAluno) AS NumEvadidos,    TotAlunos
from alunos
inner Join
    ( SELECT
        AluEscola,    count(MatricAluno) AS TotAlunos
    FROM alunos
    GROUP BY 'AluEscola '
    ) as total
    On total,AluEscola = alunos,AluEscola
WHERE ForSaidaOpcao = 2
GROUP BY 'AluEscola '

SELECT
    alunos,AluEscola, MotivoFormaSaidaOpcao, count(MatricAluno) AS
        NumEvadidos
from alunos
WHERE ForSaidaOpcao = 2
GROUP BY 'AluEscola ', alunos,MotivoFormaSaidaOpcao
Order by alunos,'AluEscola 'DESC

/*FORMA DE INGRESSO*/
SELECT
    alunos,ForIngressoOpcao,    count(MatricAluno) AS NumEvadidos,    TotAlunos
from alunos
inner Join
    ( SELECT
        ForIngressoOpcao,    count(MatricAluno) AS TotAlunos
    FROM alunos
    GROUP BY 'ForIngressoOpcao '
    ) as total
    On total,ForIngressoOpcao = alunos,ForIngressoOpcao

```

```

WHERE ForSaidaOpcao = 2
GROUP BY 'ForIngressoOpcao '

/*DESEMPENHO*/
Select
case when 'Desempenho' = 0 then
    '0'
    when 'Desempenho' > 0 and 'Desempenho' < 0.3 then
        '0,3'
    when 'Desempenho' >= 0.3 and 'Desempenho' < 0.5 then
        '0,3'
    when 'Desempenho' >= 0.5 and 'Desempenho' < 0.7 then
        '0,5'
    when 'Desempenho' >= 0.7 and 'Desempenho' < 0.9 then
        '0,7'
    when 'Desempenho' >= 0.9 and 'Desempenho' <= 1 then
        '0,9'
    end as faixa_desempenho
, count(*) as Quantidade
From alunos
Where ForSaidaOpcao = 2
Group by faixa_desempenho

/*MEDIA DE CREDITOS POR SEMESTRE*/
SELECT
historico.SemetreCursado,
    sum(historico.Creditos)/count(distinct(historico.'MatricAluno')) as
    M dia_de_creditos
FROM alunos inner join historico
on alunos.'MatricAluno'=historico.'MatricAluno '
where alunos.GrupoAnalise = 0
group by historico.SemetreCursado

SELECT
historico.SemetreCursado,
    sum(historico.Creditos)/count(distinct(historico.'MatricAluno')) as
    M dia_de_creditos
FROM alunos inner join historico
on alunos.'MatricAluno'=historico.'MatricAluno '
JOIN obrigatorias ON obrigatorias.CodDisciplina = historico.Disciplina
where alunos.GrupoAnalise = 0 and alunos.FormaSaida=2
group by historico.SemetreCursado

SELECT
historico.SemetreCursado,
    sum(historico.Creditos)/count(distinct(historico.'MatricAluno')) as
    M dia_de_creditos
FROM alunos inner join historico
on alunos.'MatricAluno'=historico.'MatricAluno '
JOIN obrigatorias ON obrigatorias.CodDisciplina = historico.Disciplina
where alunos.GrupoAnalise = 0 and alunos.FormaSaida=1
group by historico.SemetreCursado

/*TAXA DE REPROVACAO*/
Select
case when 'TaxaReprovacao' = 0 then

```

```

        '0'
        when 'TaxaReprovacao'>0 and 'TaxaReprovacao'<0.3 then
        '0,3<=Tx,Reprova o<=0,3'
    when 'TaxaReprovacao'>=0.3 and 'TaxaReprovacao'<0.5 then
        '0,3<=Tx,Reprova o<=0,5'
    when 'TaxaReprovacao'>=0.5 and 'TaxaReprovacao'<0.7 then
        '0,5<=Tx,Reprova o<=0,7'
    when 'TaxaReprovacao'>=0.7 and 'TaxaReprovacao'<0.9 then
        '0,7<=Tx,Reprova o<=0,9'
    when 'TaxaReprovacao'>=0.9 and 'TaxaReprovacao'<=1 then
        '0,9<=Tx,Reprova o<=1'
    end as faixa_TaxaReprovacao
, count(*) as Quantidade
From alunos
Where ForSaidaOpcao = 2
Group by faixa_TaxaReprovacao

/*MENCOES NAS DISCIPLINAS*/
SELECT
historico.Mencao, count(historico.MatricAluno) as Quantidade_de_alunos ,
100*count(historico.MatricAluno)/(select count(historico.MatricAluno)from
    alunos inner join historico
on alunos.MatricAluno=historico.MatricAluno where alunos.GrupoAnalise = 0
    and historico.Disciplina=118001 and alunos.FormaSaida=2) as Porcentagem
FROM alunos inner join historico
on alunos.MatricAluno=historico.MatricAluno
where alunos.GrupoAnalise = 0 and historico.Disciplina=118001 and
    alunos.FormaSaida=2
group by historico.Mencao
Order by field(historico.Mencao, 'TR', 'TJ', 'SR', 'II', 'MI', 'MM',
    'MS', 'SS', 'CC', 'AP', 'DP')

Select obrigatorias.CodDisciplina ,count(historico.MatricAluno) as
Reprovados
From alunos inner join historico
on alunos.MatricAluno=historico.MatricAluno
join obrigatorias on historico.Disciplina=obrigatorias.CodDisciplina
where alunos.GrupoAnalise = 0 And historico.Aprovacao = 0 and
    obrigatorias.Fluxo = 1
group by obrigatorias.CodDisciplina

Select historico.AnoDisciplina ,count(historico.MatricAluno) as Reprovados
From alunos inner join historico
on alunos.MatricAluno=historico.MatricAluno
where alunos.GrupoAnalise = 0 And historico.Aprovacao = 0 and
    historico.Disciplina=118001
group by historico.AnoDisciplina

/*REPROVACAO NAS MATERIAS*/
SELECT count(MatricAluno) FROM historico WHERE Disciplina = 116301

SELECT count(MatricAluno) FROM historico WHERE Disciplina = 116301 AND
    Aprovacao = 0

```

Apêndice C

Dados da análise estatística

Tabela C.1: Índice de evasão por semestre e por motivo de saída. O código do motivo de saída está de acordo com o apresentado na Tabela 5.3. O semestre 0 corresponde a um semestre cursado no verão, independente da posição do fluxo em que o aluno estava quando cursou.

Semestre	Motivo de Saída	Quantidade de Alunos	%
0	7	3	0.6383
0	21	1	0.2128
0	20	1	0.2128
1	7	11	2.3404
1	5	4	0.8511
1	21	2	0.4255
1	17	2	0.4255
2	16	22	4.6809
2	7	11	2.3404
2	21	8	1.7021
2	5	7	1.4894
2	12	2	0.4255
2	17	1	0.2128
2	20	1	0.2128
3	17	78	16.5957
3	20	14	2.9787
3	7	13	2.766
3	21	9	1.9149
3	6	5	1.0638
3	16	2	0.4255
3	12	1	0.2128
3	5	1	0.2128
3	52	1	0.2128
4	17	26	5.5319

Continua na página seguinte

Tabela C.1 – *Continuação da tabela da página anterior*

Semestre	Motivo de Saída	Quantidade de Alunos	%
4	21	10	2.1277
4	16	8	1.7021
4	20	5	1.0638
4	7	4	0.8511
4	52	3	0.6383
4	6	2	0.4255
5	17	26	5.5319
5	16	7	1.4894
5	21	7	1.4894
5	7	6	1.2766
5	52	3	0.6383
5	20	3	0.6383
6	17	20	4.2553
6	20	6	1.2766
6	16	6	1.2766
6	7	5	1.0638
6	21	4	0.8511
7	17	14	2.9787
7	7	4	0.8511
7	21	3	0.6383
7	16	2	0.4255
7	52	2	0.4255
8	17	9	1.9149
8	20	3	0.6383
8	7	3	0.6383
8	16	2	0.4255
8	6	1	0.2128
9	17	4	0.8511
9	7	3	0.6383
9	21	2	0.4255
9	16	2	0.4255
9	20	1	0.2128
9	52	1	0.2128
10	17	11	2.3404
10	20	3	0.6383
10	16	3	0.6383
10	7	1	0.2128
11	20	4	0.8511
11	16	3	0.6383
11	17	2	0.4255
12	16	7	1.4894
12	7	3	0.6383

Continua na página seguinte

Tabela C.1 – *Continuação da tabela da página anterior*

Semestre	Motivo de Saída	Quantidade de Alunos	%
12	17	3	0.6383
12	21	1	0.2128
12	20	1	0.2128
12	52	1	0.2128
13	17	3	0.6383
13	16	2	0.4255
14	16	4	0.8511
14	21	1	0.2128
14	17	1	0.2128
15	17	3	0.6383
15	16	1	0.2128
16	17	2	0.4255
16	16	1	0.2128
18	21	1	0.2128
18	17	1	0.2128

Tabela C.2: Índice de formatura por semestre. O semestre 0 corresponde a um semestre cursado no verão, independente da posição do fluxo em que o aluno estava quando cursou.

Semestre	Quantidade de Alunos
0	8
1	1
2	2
3	2
4	1
5	3
6	6
7	8
8	11
9	70
10	125
11	59
12	17
13	22
14	16
15	14
16	5
17	1
20	2

Tabela C.3: Taxa de evasão feminina e motivos de saída. O código do motivo de saída está de acordo com o apresentado na Tabela 5.3.

Motivo de saída	Quantidade de alunos	%
5	2	3.85
6	1	1.92
7	11	21.15
12	3	5.77
16	11	21.15
17	11	21.15
20	1	1.92
21	8	15.39
52	4	7.7
Total	52	100.0

Tabela C.4: Taxa de evasão masculina e motivos de saída. O código do motivo de saída está de acordo com o apresentado na Tabela 5.3.

Motivo de saída	Quantidade de alunos	%
5	10	2.39
6	7	1.67
7	56	13.4
16	61	14.6
17	195	46.65
20	41	9.81
21	41	9.81
52	7	1.67
Total	418	100.0

Tabela C.5: Índice de evasão por idade e por motivo de saída. O código do motivo de saída está de acordo com o apresentado na Tabela 5.3.

Idade	Motivo de Saída	Quantidade de Alunos	%
17	16	3	0.6383
17	17	1	0.2128
18	7	12	2.5532
18	16	4	0.8511
18	17	4	0.8511
18	21	4	0.8511

Continua na página seguinte

Tabela C.5 – *Continuação da tabela da página anterior*

Idade	Motivo de Saída	Quantidade de Alunos	%
18	52	1	0.2128
18	6	1	0.2128
19	17	30	6.383
19	21	14	2.9787
19	7	9	1.9149
19	16	6	1.2766
19	20	5	1.0638
19	6	1	0.2128
19	52	1	0.2128
20	17	28	5.9574
20	7	11	2.3404
20	20	10	2.1277
20	21	8	1.7021
20	16	7	1.4894
20	5	2	0.4255
20	6	2	0.4255
20	52	2	0.4255
21	17	25	5.3191
21	21	9	1.9149
21	7	8	1.7021
21	16	5	1.0638
21	20	5	1.0638
21	52	4	0.8511
21	6	2	0.4255
21	5	1	0.2128
22	17	29	6.1702
22	7	14	2.9787
22	20	10	2.1277
22	16	7	1.4894
22	21	4	0.8511
22	5	1	0.2128
23	17	18	3.8298
23	20	5	1.0638
23	7	4	0.8511
23	16	3	0.6383
23	21	2	0.4255
23	5	2	0.4255
23	52	1	0.2128
23	6	1	0.2128
24	17	18	3.8298
24	16	8	1.7021
24	20	2	0.4255

Continua na página seguinte

Tabela C.5 – *Continuação da tabela da página anterior*

Idade	Motivo de Saída	Quantidade de Alunos	%
24	7	2	0.4255
24	52	1	0.2128
24	12	1	0.2128
24	6	1	0.2128
25	16	12	2.5532
25	17	11	2.3404
25	21	3	0.6383
25	7	3	0.6383
25	20	3	0.6383
25	5	1	0.2128
26	17	11	2.3404
26	5	3	0.6383
26	16	3	0.6383
26	21	1	0.2128
27	17	4	0.8511
27	16	3	0.6383
27	21	3	0.6383
27	7	2	0.4255
27	5	2	0.4255
28	17	5	1.0638
28	16	3	0.6383
28	20	1	0.2128
28	7	1	0.2128
29	17	3	0.6383
29	16	1	0.2128
29	7	1	0.2128
29	20	1	0.2128
30	17	5	1.0638
31	17	4	0.8511
31	16	2	0.4255
31	12	1	0.2128
32	17	3	0.6383
32	16	1	0.2128
33	17	2	0.4255
34	17	1	0.2128
36	17	1	0.2128
37	16	1	0.2128
37	21	1	0.2128
38	16	3	0.6383
38	17	1	0.2128
39	17	1	0.2128
39	12	1	0.2128

Continua na página seguinte

Tabela C.5 – *Continuação da tabela da página anterior*

Idade	Motivo de Saída	Quantidade de Alunos	%
42	17	1	0.2128
44	52	1	0.2128

Tabela C.6: Índice de formatura por idade.

Idade	Quantidade de Alunos	%
20	1	0.2128
21	15	3.1915
22	77	16.383
23	100	21.2766
24	68	14.4681
25	41	8.7234
26-30	62	13.1915
>30	9	1.9151

Tabela C.7: Taxa de evasão por tipo de escola e motivos de saída. O código do motivo de saída está de acordo com o apresentado na Tabela 5.3.

Tipo de escola	Motivo de saída	Quantidade de Alunos
Não declarado	5	9
Não declarado	6	6
Não declarado	7	61
Não declarado	12	3
Não declarado	16	58
Não declarado	17	130
Não declarado	20	22
Não declarado	21	19
Não declarado	52	4
Pública	5	3
Pública	6	1
Pública	7	3
Pública	16	5
Pública	17	19
Pública	20	6
Pública	21	7
Pública	52	2
Particular	6	1
Particular	7	3
Particular	16	9
Particular	17	57
Particular	20	14
Particular	21	23
Particular	52	5

Tabela C.8: Taxa de evasão por forma de ingresso. O código da forma de ingresso está de acordo com o apresentado na Tabela 5.6. A quantidade de evadidos representa quantos alunos que entraram por cada uma das formas de ingresso e evadiram. O total de alunos é a quantidade de alunos que entraram por aquela forma de ingresso, independente da situação atual dele no curso, ou seja, se está cursando, formado ou evadido.

Forma de ingresso	Quantidade de evadidos	Total de alunos	% de evadidos
1	259	706	36.68555241
15	29	47	61.70212766
17	81	240	33.75
2	43	70	61.42857143
20	1	1	100
24	4	4	100
3	11	27	40.74074074
5	5	9	55.55555556
52	2	22	9.090909091
6	9	9	100
7	26	27	96.2962963

Tabela C.9: Desempenho médio dos evadidos organizados em faixas de desempenho.

Desempenho	Número de evadidos
0	52
$0 < \text{Desempenho} < 0,3$	80
$0,3 \leq \text{Desempenho} < 0,5$	101
$0,5 \leq \text{Desempenho} < 0,7$	129
$0,7 \leq \text{Desempenho} < 0,9$	67
$0,9 \leq \text{Desempenho} \leq 1$	41

Tabela C.10: Desempenho médio dos formados organizados em faixas de desempenho.

Desempenho	Número de formados
$0,5 \leq \text{Desempenho} < 0,7$	19
$0,7 \leq \text{Desempenho} < 0,9$	145
$0,9 \leq \text{Desempenho} \leq 1$	209

Tabela C.11: Taxa de reprovação média dos evadidos organizados em faixas de reprovação.

Desempenho	Número de evadidos
0	44
$0 < \text{Desempenho} < 0,3$	85
$0,3 \leq \text{Desempenho} < 0,5$	121
$0,5 \leq \text{Desempenho} < 0,7$	109
$0,7 \leq \text{Desempenho} < 0,9$	67
$0,9 \leq \text{Desempenho} \leq 1$	44

Tabela C.12: Taxa de reprovação média dos formados organizados em faixas de reprovação.

Desempenho	Número de formados
0	81
$0 < \text{Desempenho} < 0,3$	273
$0,3 \leq \text{Desempenho} < 0,5$	19

Tabela C.13: Média de créditos obrigatórios cursados por semestre dos evadidos. O semestre 0 representa um semestre cursado no verão, independente do período curricular que ele estava quando cursou.

Semestre	Média de créditos
0	8.50
1	24.22
2	18.65
3	17.69
4	17.43
5	18.02
6	17.11
7	16.43
8	14.74
9	14.15
10	12.55
11	14.38
12	9.71
13	8.29
14	13.60
15	3.33
16	4.00
17	4.00
18	4.00

Tabela C.14: Média de créditos obrigatórios cursados por semestre dos formados. O semestre 0 representa um semestre cursado no verão, independente do período curricular que ele estava quando cursou.

Semestre	Média de créditos
0	9.56
1	26.14
2	22.54
3	22.43
4	20.91
5	22.51
6	18.97
7	16.08
8	11.52
9	9.20
10	8.64
11	8.81
12	9.47
13	8.35
14	6.88
15	6.80
16	7.33
17	8.00
18	4.00
19	12.00
20	8.00

Tabela C.15: Índice de reprovação das matérias obrigatórias organizadas por departamento dos alunos evadidos e dos formados.

Departamento	Evadido	Formado
CIC	0.49	0.07
EST	0.20	0,06
MAT	0.59	0.16
IFD	0.47	0.06
LIP	0.12	0.01
LET	0.12	0.03
FIL	0.08	0.05

Tabela C.16: Reprovações nas disciplinas obrigatórias recomendadas de serem cursadas no 1º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10.

Disciplina	Taxa de reprovação
116301	29.72973
113034	39.13669
118001	40.8998
118010	30.50975

Tabela C.17: Reprovações nas disciplinas obrigatórias recomendadas de serem cursadas no 2º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10.

Disciplina	Taxa de reprovação
113042	32.21071
116319	24.43845
118028	26.40509
118036	14.28571

Tabela C.18: Reprovações nas disciplinas obrigatórias recomendadas de serem cursadas no 3º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10.

Disciplina	Taxa de reprovação
113051	42.91225
118044	31.80212
118052	13.32418
113956	22.60274
117366	34.0591

Tabela C.19: Reprovações nas disciplinas obrigatórias recomendadas de serem cursadas no 4º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10.

Disciplina	Taxa de reprovação
113123	29.39597
116351	16.74208
116416	8.020699
116327	15.45338

Tabela C.20: Reprovações nas disciplinas obrigatórias recomendadas de serem cursadas no 5º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10.

Disciplina	Taxa de reprovação
113107	39.1657
116378	11.61388
113417	31.20125
116343	17.79891
116394	25.11211
113115	27.57794

Tabela C.21: Evolução da taxa de reprovação ao longo dos anos nas disciplinas obrigatórias recomendadas de serem cursadas no 1º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10.

Ano	113034	116301	118001	118010
2000	25.71	17.14	33.82	20.00
2001	23.81	13.75	20.65	22.62
2002	21.11	24.42	10.84	20.45
2003	27.72	21.90	23.40	15.63
2004	24.14	22.22	16.05	27.91
2005	41.90	16.48	22.34	10.00
2006	34.95	27.66	25.00	26.14
2007	33.80	23.08	21.33	15.00
2008	25.68	23.38	33.33	17.91
2009	45.79	36.36	40.20	25.88
2010	40.00	33.09	57.78	35.11
2011	59.69	40.16	53.59	47.83
2012	58.73	46.15	69.51	46.56
2013	55.08	47.58	69.88	60.94

Tabela C.22: Evolução da taxa de reprovação ao longo dos anos nas disciplinas obrigatórias recomendadas de serem cursadas no 2º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10.

Ano	113042	118028
2000	11.11	8.70
2001	1.59	21.05
2002	15.38	17.07
2003	37.04	14.46
2004	26.80	10.77
2005	13.58	6.35
2006	32.26	19.48
2007	23.47	25.30
2008	36.76	20.75
2009	32.88	14.29
2010	48.15	52.63
2011	44.44	34.02
2012	45.28	43.24
2013	50.53	63.93

Tabela C.23: Evolução da taxa de reprovação ao longo dos anos nas disciplinas obrigatórias recomendadas de serem cursadas no 3º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10.

Ano	113051	117366	118044
2000	0.00	-	50.00
2001	50.00	-	16.22
2002	50.00	-	13.64
2003	53.85	-	20.29
2004	45.87	-	27.14
2005	44.58	23.33	30.38
2006	35.19	39.36	30.00
2007	38.71	12.07	38.75
2008	40.00	44.59	36.99
2009	50.00	51.61	47.27
2010	37.50	16.67	33.33
2011	32.81	32.00	49.09
2012	45.76	39.78	38.96
2013	42.62	31.19	27.12

Tabela C.24: Evolução da taxa de reprovação ao longo dos anos nas disciplinas obrigatórias recomendadas de serem cursadas no 4º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10.

Ano	113123
2000	0.00
2001	20.00
2002	27.91
2003	40.00
2004	26.56
2005	37.50
2006	36.49
2007	25.76
2008	13.24
2009	18.33
2010	13.95
2011	40.43
2012	21.31
2013	72.22

Tabela C.25: Evolução da taxa de reprovação ao longo dos anos nas disciplinas obrigatórias recomendadas de serem cursadas no 4º semestre do curso. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10.

Ano	113107	113115	113417	116394
2000	20.00	0.00	0.00	20.00
2001	41.67	25.00	30.77	33.33
2002	30.95	12.00	29.41	12.82
2003	23.08	21.31	12.31	24.56
2004	38.57	15.00	23.73	10.34
2005	33.33	13.58	39.68	10.00
2006	30.86	26.56	39.51	19.61
2007	36.92	32.84	28.05	22.81
2008	46.81	29.73	36.36	30.00
2009	32.56	22.83	27.59	24.05
2010	43.75	42.22	24.24	37.04
2011	42.00	36.36	31.25	29.41
2012	39.66	23.61	27.03	34.21
2013	70.83	57.35	64.86	48.15

Tabela C.26: Menções obtidas pelos alunos evadidos nas disciplinas obrigatórias. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10.

Disciplina	TR	TJ	SR	II	MI	MM	MS	SS	CC
116301	1.45	0.21	12.19	17.15	19.42	20.66	17.36	5.99	5.58
113034	1.47	0.21	19.50	24.95	16.56	21.80	9.01	1.26	5.24
118001	2.96	2.37	16.57	22.88	15.38	25.84	8.88	3.94	1.18
118010	2.62	1.09	14.63	8.52	21.18	27.95	18.78	4.15	1.09
113042	0.68	1.36	15.31	18.71	17.35	32.31	9.86	2.04	2.38
118028	1.83	0.00	13.70	9.13	20.09	38.81	14.61	1.83	0.00
113051	7.91	0.00	24.86	19.21	15.82	20.90	8.47	1.13	1.69
118044	4.19	0.60	25.75	10.78	20.36	28.74	7.78	0.00	1.80
117366	6.25	0.78	32.81	10.16	11.72	19.53	11.72	2.34	4.69
113123	19.32	0.00	14.77	11.36	13.64	27.27	10.23	2.27	1.14
113107	6.25	0.00	30.47	22.66	7.81	21.88	7.03	0.00	3.91
113417	5.80	0.00	5.80	23.19	15.94	17.39	24.64	7.25	5.80
116394	1.47	0.00	20.59	11.76	17.65	17.65	11.76	2.94	16.18
113115	2.94	0.00	19.85	12.50	20.59	30.88	10.29	2.21	0.74

Tabela C.27: Menções obtidas pelos alunos formados nas disciplinas obrigatórias. Os códigos das disciplinas estão de acordo com os apresentados nas Tabelas 2.5, 2.7, 2.8, 2.9 e 2.10.

Disciplina	TR	TJ	SR	II	MI	MM	MS	SS	CC
116301	0.00	0.00	1.20	0.60	3.30	22.22	39.34	30.33	3.00
113034	0.29	0.00	0.58	2.05	9.06	42.98	28.95	11.11	4.97
118001	0.00	0.00	2.06	1.18	5.00	37.35	32.94	18.53	2.94
118010	0.00	0.00	0.60	0.60	5.95	31.55	46.13	13.99	1.19
113042	1.42	0.00	0.57	3.70	5.98	44.44	29.34	13.11	1.42
118028	1.83	0.00	13.70	9.13	20.09	38.81	14.61	1.83	0.00
113051	4.35	0.00	3.93	12.01	15.32	44.31	16.56	2.69	0.83
118044	0.00	0.00	1.48	3.85	9.47	52.66	21.89	9.76	0.89
117366	1.89	0.00	5.03	1.26	13.84	35.22	28.30	13.84	0.63
113123	5.53	0.00	1.58	4.47	6.58	45.53	23.68	12.11	0.53
113107	2.05	0.00	3.33	6.92	11.28	41.79	26.41	7.69	0.51
113417	3.14	0.00	2.00	4.29	10.86	39.43	29.71	9.43	1.14
116394	0.00	0.00	2.46	3.28	8.74	41.80	35.25	6.01	2.46
113115	0.83	0.00	1.11	1.93	9.67	41.71	30.66	13.81	0.28

Apêndice D

Consulta SQL para mineração de dados

A seguinte consulta SQL foi utilizada para criação da tabela do banco de dados utilizado para mineração de dados.

```
CREATE TABLE mineracao_departamentos_cursando LIKE alunos;
INSERT mineracao_departamentos_cursando SELECT * FROM alunos;

ALTER TABLE 'mineracao_departamentos_cursando'
DROP 'AnoIngresso',
DROP 'SemestreIngresso',
DROP 'FormaIngresso',
DROP 'AnoSaida',
DROP 'SemestreSaida',
DROP 'FormaSaida',
DROP 'PerIngressoOpcao',
DROP 'SemestreIngressoOpcao',
DROP 'PerSaidaOpcao',
DROP 'SemestreSaidaOpcao',
DROP 'AlunoRegistrado',
DROP 'PeriodoCurricular',
DROP 'SemestrePeriodoCurricular',
DROP 'AluNacionalidade',
DROP 'AluDtNasc',
DROP 'AluCotId',
DROP 'AluEscola',
DROP 'AnoNasc',
DROP 'IdadeSaida',
DROP 'SemestreCursoSaida',
DROP 'Desempenho',
DROP 'TaxaReprovacao',
DROP 'MotivoFormaSaida',
DROP 'MotivoFormaSaidaOpcao',
DROP 'AnoSemetreIngresso',
DROP 'AnoSemestreIngressoOpcao',
DROP 'AnoSemestreSaida',
DROP 'AnoSemestreSaidaOpcao';

DELETE FROM 'mineracao_departamentos_cursando' WHERE 'ForSaidaOpcao'=0;
DELETE FROM 'mineracao_departamentos_cursando' WHERE 'GrupoAnalise'=1;
DELETE FROM 'mineracao_departamentos_cursando' WHERE 'GrupoAnalise'=2;

ALTER TABLE 'mineracao_departamentos_cursando'
```

```

DROP 'GrupoAnalise';

ALTER TABLE 'mineracao_departamentos_cursando' ADD 'IRAobr' FLOAT NOT NULL
;
UPDATE mineracao_departamentos_cursando INNER JOIN historico ON
mineracao_departamentos_cursando.'MatricAluno'=historico.'MatricAluno'
SET
mineracao_departamentos_cursando.'IRAobr'=historico.IndiceRendimentoObr;

UPDATE 'mineracao_departamentos_cursando' SET 'IRAobr'=
CASE WHEN 'IRAobr' = 0 THEN '6'
WHEN 'IRAobr'>0 AND 'IRAobr'<1 THEN'7'
WHEN 'IRAobr'>=1 AND 'IRAobr'<2 THEN'8'
WHEN 'IRAobr'>=2 AND 'IRAobr'<3 THEN'9'
WHEN 'IRAobr'>=3 AND 'IRAobr'<4 THEN'10'
WHEN 'IRAobr'>=4 AND 'IRAobr'<=5 THEN'11'
end;

ALTER TABLE 'mineracao_departamentos_cursando' ADD 'Cic' FLOAT NOT NULL ,
ADD 'Est' FLOAT NOT NULL , ADD 'Mat' FLOAT NOT NULL , ADD 'IFD'
FLOAT NOT NULL , ADD 'Lip' FLOAT NOT NULL , ADD 'Let' FLOAT NOT NULL
, ADD 'Fil' FLOAT NOT NULL ;

UPDATE 'mineracao_departamentos_cursando' INNER JOIN
(SELECT reprovadas.MatricAluno, reprovadas.creditosRepro/total.creditosTot
AS taxa FROM

(SELECT alunos.MatricAluno, SUM(historico.creditos) AS creditosTot FROM
alunos INNER JOIN historico ON alunos.MatricAluno =
historico.MatricAluno JOIN obrigatorias ON obrigatorias.CodDisciplina =
historico.Disciplina WHERE obrigatorias.Departamento='CIC' GROUP BY
alunos.MatricAluno) AS total
INNER JOIN
(SELECT alunos.MatricAluno, SUM(historico.creditos) AS creditosRepro FROM
alunos INNER JOIN historico ON alunos.MatricAluno =
historico.MatricAluno JOIN obrigatorias ON obrigatorias.CodDisciplina =
historico.Disciplina WHERE historico.Aprovacao = 0 AND
obrigatorias.Departamento='CIC' GROUP BY alunos.MatricAluno) AS
reprovadas

ON reprovadas.MatricAluno=total.MatricAluno
GROUP BY reprovadas.MatricAluno) AS ReprovacaoAlunos

ON
ReprovacaoAlunos.MatricAluno='mineracao_departamentos_cursando'.MatricAluno

SET 'mineracao_departamentos_cursando'.Cic=ReprovacaoAlunos.taxa;

UPDATE 'mineracao_departamentos_cursando' INNER JOIN
(SELECT reprovadas.MatricAluno, reprovadas.creditosRepro/total.creditosTot
AS taxa FROM

(SELECT alunos.MatricAluno, SUM(historico.creditos) AS creditosTot FROM
alunos INNER JOIN historico ON alunos.MatricAluno =

```

```

historico.MatricAluno JOIN obrigatorias ON obrigatorias.CodDisciplina =
historico.Disciplina WHERE obrigatorias.Departamento='EST' GROUP BY
alunos.MatricAluno) AS total
INNER JOIN
(SELECT alunos.MatricAluno, SUM(historico.creditos) AS creditosRepro FROM
alunos INNER JOIN historico ON alunos.MatricAluno =
historico.MatricAluno JOIN obrigatorias ON obrigatorias.CodDisciplina =
historico.Disciplina WHERE historico.Aprovacao = 0 AND
obrigatorias.Departamento='EST' GROUP BY alunos.MatricAluno) AS
reprovadas

ON reprovadas.MatricAluno=total.MatricAluno
GROUP BY reprovadas.MatricAluno) AS ReprovacaoAlunos

ON
ReprovacaoAlunos.MatricAluno='mineracao_departamentos_cursando'.MatricAluno

SET 'mineracao_departamentos_cursando'.Est=ReprovacaoAlunos.taxa;

UPDATE 'mineracao_departamentos_cursando' INNER JOIN
(SELECT reprovadas.MatricAluno, reprovadas.creditosRepro/total.creditosTot
AS taxa FROM

(SELECT alunos.MatricAluno, SUM(historico.creditos) AS creditosTot FROM
alunos INNER JOIN historico ON alunos.MatricAluno =
historico.MatricAluno JOIN obrigatorias ON obrigatorias.CodDisciplina =
historico.Disciplina WHERE obrigatorias.Departamento='MAT' GROUP BY
alunos.MatricAluno) AS total
INNER JOIN
(SELECT alunos.MatricAluno, SUM(historico.creditos) AS creditosRepro FROM
alunos INNER JOIN historico ON alunos.MatricAluno =
historico.MatricAluno JOIN obrigatorias ON obrigatorias.CodDisciplina =
historico.Disciplina WHERE historico.Aprovacao = 0 AND
obrigatorias.Departamento='MAT' GROUP BY alunos.MatricAluno) AS
reprovadas

ON reprovadas.MatricAluno=total.MatricAluno
GROUP BY reprovadas.MatricAluno) AS ReprovacaoAlunos

ON
ReprovacaoAlunos.MatricAluno='mineracao_departamentos_cursando'.MatricAluno

SET 'mineracao_departamentos_cursando'.Mat=ReprovacaoAlunos.taxa;

UPDATE 'mineracao_departamentos_cursando' INNER JOIN
(SELECT reprovadas.MatricAluno, reprovadas.creditosRepro/total.creditosTot
AS taxa FROM

(SELECT alunos.MatricAluno, SUM(historico.creditos) AS creditosTot FROM
alunos INNER JOIN historico ON alunos.MatricAluno =
historico.MatricAluno JOIN obrigatorias ON obrigatorias.CodDisciplina =
historico.Disciplina WHERE obrigatorias.Departamento='IFD' GROUP BY
alunos.MatricAluno) AS total

```



```

INNER JOIN
(SELECT alunos.MatricAluno, SUM(historico.creditos) AS creditosRepro FROM
alunos INNER JOIN historico ON alunos.MatricAluno =
historico.MatricAluno JOIN obrigatorias ON obrigatorias.CodDisciplina =
historico.Disciplina WHERE historico.Aprovacao = 0 AND
obrigatorias.Departamento='IFD' GROUP BY alunos.MatricAluno) AS
reprovadas

ON reprovadas.MatricAluno=total.MatricAluno
GROUP BY reprovadas.MatricAluno) AS ReprovacaoAlunos

ON
ReprovacaoAlunos.MatricAluno='mineracao_departamentos_cursando'.MatricAluno

SET 'mineracao_departamentos_cursando'.IFD=ReprovacaoAlunos.taxa;

UPDATE 'mineracao_departamentos_cursando' INNER JOIN
(SELECT reprovadas.MatricAluno, reprovadas.creditosRepro/total.creditosTot
AS taxa FROM

(SELECT alunos.MatricAluno, SUM(historico.creditos) AS creditosTot FROM
alunos INNER JOIN historico ON alunos.MatricAluno =
historico.MatricAluno JOIN obrigatorias ON obrigatorias.CodDisciplina =
historico.Disciplina WHERE obrigatorias.Departamento='LIP' GROUP BY
alunos.MatricAluno) AS total
INNER JOIN
(SELECT alunos.MatricAluno, SUM(historico.creditos) AS creditosRepro FROM
alunos INNER JOIN historico ON alunos.MatricAluno =
historico.MatricAluno JOIN obrigatorias ON obrigatorias.CodDisciplina =
historico.Disciplina WHERE historico.Aprovacao = 0 AND
obrigatorias.Departamento='LIP' GROUP BY alunos.MatricAluno) AS
reprovadas

ON reprovadas.MatricAluno=total.MatricAluno
GROUP BY reprovadas.MatricAluno) AS ReprovacaoAlunos

ON
ReprovacaoAlunos.MatricAluno='mineracao_departamentos_cursando'.MatricAluno

SET 'mineracao_departamentos_cursando'.Lip=ReprovacaoAlunos.taxa;

UPDATE 'mineracao_departamentos_cursando' INNER JOIN
(SELECT reprovadas.MatricAluno, reprovadas.creditosRepro/total.creditosTot
AS taxa FROM

(SELECT alunos.MatricAluno, SUM(historico.creditos) AS creditosTot FROM
alunos INNER JOIN historico ON alunos.MatricAluno =
historico.MatricAluno JOIN obrigatorias ON obrigatorias.CodDisciplina =
historico.Disciplina WHERE obrigatorias.Departamento='LET' GROUP BY
alunos.MatricAluno) AS total
INNER JOIN
(SELECT alunos.MatricAluno, SUM(historico.creditos) AS creditosRepro FROM
alunos INNER JOIN historico ON alunos.MatricAluno =
historico.MatricAluno JOIN obrigatorias ON obrigatorias.CodDisciplina =

```

```

historico.Disciplina WHERE historico.Aprovacao = 0 AND
obrigatorias.Departamento='LET' GROUP BY alunos.MatricAluno) AS
reprovadas

ON reprovadas.MatricAluno=total.MatricAluno
GROUP BY reprovadas.MatricAluno) AS ReprovacaoAlunos

ON
ReprovacaoAlunos.MatricAluno='mineracao_departamentos_cursando'.MatricAluno

SET 'mineracao_departamentos_cursando'.Let=ReprovacaoAlunos.taxa;

UPDATE 'mineracao_departamentos_cursando' INNER JOIN
(SELECT reprovadas.MatricAluno, reprovadas.creditosRepro/total.creditosTot
AS taxa FROM

(SELECT alunos.MatricAluno, SUM(historico.creditos) AS creditosTot FROM
alunos INNER JOIN historico ON alunos.MatricAluno =
historico.MatricAluno JOIN obrigatorias ON obrigatorias.CodDisciplina =
historico.Disciplina WHERE obrigatorias.Departamento='Fil' GROUP BY
alunos.MatricAluno) AS total
INNER JOIN
(SELECT alunos.MatricAluno, SUM(historico.creditos) AS creditosRepro FROM
alunos INNER JOIN historico ON alunos.MatricAluno =
historico.MatricAluno JOIN obrigatorias ON obrigatorias.CodDisciplina =
historico.Disciplina WHERE historico.Aprovacao = 0 AND
obrigatorias.Departamento='Fil' GROUP BY alunos.MatricAluno) AS
reprovadas

ON reprovadas.MatricAluno=total.MatricAluno
GROUP BY reprovadas.MatricAluno) AS ReprovacaoAlunos

ON
ReprovacaoAlunos.MatricAluno='mineracao_departamentos_cursando'.MatricAluno

SET 'mineracao_departamentos_cursando'.Fil=ReprovacaoAlunos.taxa;

UPDATE 'mineracao_departamentos_cursando' SET 'Cic' =
CASE WHEN 'Cic' = 0 THEN '10'
WHEN 'Cic' > 0 AND 'Cic' < 0.1 THEN '11'
WHEN 'Cic' >= 0.1 AND 'Cic' < 0.2 THEN '12'
WHEN 'Cic' >= 0.2 AND 'Cic' < 0.3 THEN '13'
WHEN 'Cic' >= 0.3 AND 'Cic' < 0.4 THEN '14'
WHEN 'Cic' >= 0.4 AND 'Cic' < 0.5 THEN '15'
WHEN 'Cic' >= 0.5 AND 'Cic' < 0.6 THEN '16'
WHEN 'Cic' >= 0.6 AND 'Cic' < 0.7 THEN '17'
WHEN 'Cic' >= 0.7 AND 'Cic' < 0.8 THEN '18'
WHEN 'Cic' >= 0.8 AND 'Cic' < 0.9 THEN '19'
WHEN 'Cic' >= 0.9 AND 'Cic' < 1 THEN '20'
WHEN 'Cic' = 1 THEN '21'
end;

UPDATE 'mineracao_departamentos_cursando' SET 'Est' =
CASE WHEN 'Est' = 0 THEN '10'

```

```

WHEN 'Est'>0 AND 'Est'<0.1 THEN'11'
WHEN 'Est'>=0.1 AND 'Est'<0.2 THEN'12'
WHEN 'Est'>=0.2 AND 'Est'<0.3 THEN'13'
WHEN 'Est'>=0.3 AND 'Est'<0.4 THEN'14'
WHEN 'Est'>=0.4 AND 'Est'<0.5 THEN'15'
WHEN 'Est'>=0.5 AND 'Est'<0.6 THEN'16'
WHEN 'Est'>=0.6 AND 'Est'<0.7 THEN'17'
WHEN 'Est'>=0.7 AND 'Est'<0.8 THEN'18'
WHEN 'Est'>=0.8 AND 'Est'<0.9 THEN'19'
WHEN 'Est'>=0.9 AND 'Est'<1 THEN'20'
WHEN 'Est'=1 THEN'21'
end;

UPDATE 'mineracao_departamentos_cursando' SET 'Mat'=
CASE WHEN 'Mat' = 0 THEN '10'
WHEN 'Mat'>0 AND 'Mat'<0.1 THEN'11'
WHEN 'Mat'>=0.1 AND 'Mat'<0.2 THEN'12'
WHEN 'Mat'>=0.2 AND 'Mat'<0.3 THEN'13'
WHEN 'Mat'>=0.3 AND 'Mat'<0.4 THEN'14'
WHEN 'Mat'>=0.4 AND 'Mat'<0.5 THEN'15'
WHEN 'Mat'>=0.5 AND 'Mat'<0.6 THEN'16'
WHEN 'Mat'>=0.6 AND 'Mat'<0.7 THEN'17'
WHEN 'Mat'>=0.7 AND 'Mat'<0.8 THEN'18'
WHEN 'Mat'>=0.8 AND 'Mat'<0.9 THEN'19'
WHEN 'Mat'>=0.9 AND 'Mat'<1 THEN'20'
WHEN 'Mat'=1 THEN'21'
end;

UPDATE 'mineracao_departamentos_cursando' SET 'IFD'=
CASE WHEN 'IFD' = 0 THEN '10'
WHEN 'IFD'>0 AND 'IFD'<0.1 THEN'11'
WHEN 'IFD'>=0.1 AND 'IFD'<0.2 THEN'12'
WHEN 'IFD'>=0.2 AND 'IFD'<0.3 THEN'13'
WHEN 'IFD'>=0.3 AND 'IFD'<0.4 THEN'14'
WHEN 'IFD'>=0.4 AND 'IFD'<0.5 THEN'15'
WHEN 'IFD'>=0.5 AND 'IFD'<0.6 THEN'16'
WHEN 'IFD'>=0.6 AND 'IFD'<0.7 THEN'17'
WHEN 'IFD'>=0.7 AND 'IFD'<0.8 THEN'18'
WHEN 'IFD'>=0.8 AND 'IFD'<0.9 THEN'19'
WHEN 'IFD'>=0.9 AND 'IFD'<1 THEN'20'
WHEN 'IFD'=1 THEN'21'
end;

UPDATE 'mineracao_departamentos_cursando' SET 'Lip'=
CASE WHEN 'Lip' = 0 THEN '10'
WHEN 'Lip'>0 AND 'Lip'<0.1 THEN'11'
WHEN 'Lip'>=0.1 AND 'Lip'<0.2 THEN'12'
WHEN 'Lip'>=0.2 AND 'Lip'<0.3 THEN'13'
WHEN 'Lip'>=0.3 AND 'Lip'<0.4 THEN'14'
WHEN 'Lip'>=0.4 AND 'Lip'<0.5 THEN'15'
WHEN 'Lip'>=0.5 AND 'Lip'<0.6 THEN'16'
WHEN 'Lip'>=0.6 AND 'Lip'<0.7 THEN'17'
WHEN 'Lip'>=0.7 AND 'Lip'<0.8 THEN'18'
WHEN 'Lip'>=0.8 AND 'Lip'<0.9 THEN'19'
WHEN 'Lip'>=0.9 AND 'Lip'<1 THEN'20'
WHEN 'Lip'=1 THEN'21'

```

```

end;

UPDATE 'mineracao_departamentos_cursando' SET 'Let' =
CASE WHEN 'Let' = 0 THEN '10'
      WHEN 'Let' > 0 AND 'Let' < 0.1 THEN '11'
      WHEN 'Let' >= 0.1 AND 'Let' < 0.2 THEN '12'
      WHEN 'Let' >= 0.2 AND 'Let' < 0.3 THEN '13'
      WHEN 'Let' >= 0.3 AND 'Let' < 0.4 THEN '14'
      WHEN 'Let' >= 0.4 AND 'Let' < 0.5 THEN '15'
      WHEN 'Let' >= 0.5 AND 'Let' < 0.6 THEN '16'
      WHEN 'Let' >= 0.6 AND 'Let' < 0.7 THEN '17'
      WHEN 'Let' >= 0.7 AND 'Let' < 0.8 THEN '18'
      WHEN 'Let' >= 0.8 AND 'Let' < 0.9 THEN '19'
      WHEN 'Let' >= 0.9 AND 'Let' < 1 THEN '20'
      WHEN 'Let' = 1 THEN '21'
end;

UPDATE 'mineracao_departamentos_cursando' SET 'Fil' =
CASE WHEN 'Fil' = 0 THEN '10'
      WHEN 'Fil' > 0 AND 'Fil' < 0.1 THEN '11'
      WHEN 'Fil' >= 0.1 AND 'Fil' < 0.2 THEN '12'
      WHEN 'Fil' >= 0.2 AND 'Fil' < 0.3 THEN '13'
      WHEN 'Fil' >= 0.3 AND 'Fil' < 0.4 THEN '14'
      WHEN 'Fil' >= 0.4 AND 'Fil' < 0.5 THEN '15'
      WHEN 'Fil' >= 0.5 AND 'Fil' < 0.6 THEN '16'
      WHEN 'Fil' >= 0.6 AND 'Fil' < 0.7 THEN '17'
      WHEN 'Fil' >= 0.7 AND 'Fil' < 0.8 THEN '18'
      WHEN 'Fil' >= 0.8 AND 'Fil' < 0.9 THEN '19'
      WHEN 'Fil' >= 0.9 AND 'Fil' < 1 THEN '20'
      WHEN 'Fil' = 1 THEN '21'
end;

ALTER TABLE 'mineracao_departamentos_cursando'
  DROP 'MatricAluno';

CREATE TABLE mineracao_departamentos_ex_aluno LIKE alunos;
INSERT mineracao_departamentos_ex_aluno SELECT * FROM alunos;

ALTER TABLE 'mineracao_departamentos_ex_aluno'
  DROP 'AnoIngresso',
  DROP 'SemestreIngresso',
  DROP 'FormaIngresso',
  DROP 'AnoSaida',
  DROP 'SemestreSaida',
  DROP 'FormaSaida',
  DROP 'PerIngressoOpcao',
  DROP 'SemestreIngressoOpcao',
  DROP 'PerSaidaOpcao',
  DROP 'SemestreSaidaOpcao',
  DROP 'AlunoRegistrado',
  DROP 'PeriodoCurricular',
  DROP 'SemestrePeriodoCurricular',
  DROP 'AluNacionalidade';

```

```

DROP 'AluDtNasc',
DROP 'AluCotId',
DROP 'AluEscola',
DROP 'AnoNasc',
DROP 'IdadeSaida',
DROP 'SemestreCursoSaida',
DROP 'Desempenho',
DROP 'TaxaReprovacao',
DROP 'MotivoFormaSaida',
DROP 'MotivoFormaSaidaOpcao',
DROP 'AnoSemestreIngresso',
DROP 'AnoSemestreIngressoOpcao',
DROP 'AnoSemestreSaida',
DROP 'AnoSemestreSaidaOpcao';

DELETE FROM 'mineracao_departamentos_ex_aluno' WHERE 'ForSaidaOpcao'=0;
DELETE FROM 'mineracao_departamentos_ex_aluno' WHERE 'GrupoAnalise'=1;
DELETE FROM 'mineracao_departamentos_ex_aluno' WHERE 'GrupoAnalise'=2;

ALTER TABLE 'mineracao_departamentos_ex_aluno'
    DROP 'GrupoAnalise';

ALTER TABLE 'mineracao_departamentos_ex_aluno' ADD 'IRAobr' FLOAT NOT NULL
;
UPDATE mineracao_departamentos_ex_aluno INNER JOIN historico ON
    mineracao_departamentos_ex_aluno.'MatricAluno'=historico.'MatricAluno'
SET
    mineracao_departamentos_ex_aluno.'IRAobr'=historico.IndiceRendimentoObr;

UPDATE 'mineracao_departamentos_ex_aluno' SET 'IRAobr'=
CASE WHEN 'IRAobr' = 0 THEN '6'
    WHEN 'IRAobr'>0 AND 'IRAobr'<1 THEN '7'
    WHEN 'IRAobr'>=1 AND 'IRAobr'<2 THEN '8'
    WHEN 'IRAobr'>=2 AND 'IRAobr'<3 THEN '9'
    WHEN 'IRAobr'>=3 AND 'IRAobr'<4 THEN '10'
    WHEN 'IRAobr'>=4 AND 'IRAobr'<=5 THEN '11'
end;

ALTER TABLE 'mineracao_departamentos_ex_aluno' ADD 'Cic' FLOAT NOT NULL ,
    ADD 'Est' FLOAT NOT NULL , ADD 'Mat' FLOAT NOT NULL , ADD 'IFD'
    FLOAT NOT NULL , ADD 'Lip' FLOAT NOT NULL , ADD 'Let' FLOAT NOT NULL
    , ADD 'Fil' FLOAT NOT NULL ;

UPDATE 'mineracao_departamentos_ex_aluno' INNER JOIN
(SELECT reprovadas.MatricAluno, reprovadas.creditosRepro/total.creditosTot
AS taxa FROM

(SELECT alunos.MatricAluno, SUM(historico.creditos) AS creditosTot FROM
    alunos INNER JOIN historico ON alunos.MatricAluno =
    historico.MatricAluno JOIN obrigatorias ON obrigatorias.CodDisciplina =
    historico.Disciplina WHERE obrigatorias.Departamento='CIC' GROUP BY
    alunos.MatricAluno) AS total
INNER JOIN
(SELECT alunos.MatricAluno, SUM(historico.creditos) AS creditosRepro FROM
    alunos INNER JOIN historico ON alunos.MatricAluno =

```

```

historico.MatricAluno JOIN obrigatorias ON obrigatorias.CodDisciplina =
historico.Disciplina WHERE historico.Aprovacao = 0 AND
obrigatorias.Departamento='CIC' GROUP BY alunos.MatricAluno) AS
reprovadas

ON reprovadas.MatricAluno=total.MatricAluno
GROUP BY reprovadas.MatricAluno) AS ReprovacaoAlunos

ON
ReprovacaoAlunos.MatricAluno='mineracao_departamentos_ex_aluno'.MatricAluno

SET 'mineracao_departamentos_ex_aluno'.Cic=ReprovacaoAlunos.taxa;

UPDATE 'mineracao_departamentos_ex_aluno' INNER JOIN
(SELECT reprovadas.MatricAluno, reprovadas.creditosRepro/total.creditosTot
AS taxa FROM

(SELECT alunos.MatricAluno, SUM(historico.creditos) AS creditosTot FROM
alunos INNER JOIN historico ON alunos.MatricAluno =
historico.MatricAluno JOIN obrigatorias ON obrigatorias.CodDisciplina =
historico.Disciplina WHERE obrigatorias.Departamento='EST' GROUP BY
alunos.MatricAluno) AS total
INNER JOIN
(SELECT alunos.MatricAluno, SUM(historico.creditos) AS creditosRepro FROM
alunos INNER JOIN historico ON alunos.MatricAluno =
historico.MatricAluno JOIN obrigatorias ON obrigatorias.CodDisciplina =
historico.Disciplina WHERE historico.Aprovacao = 0 AND
obrigatorias.Departamento='EST' GROUP BY alunos.MatricAluno) AS
reprovadas

ON reprovadas.MatricAluno=total.MatricAluno
GROUP BY reprovadas.MatricAluno) AS ReprovacaoAlunos

ON
ReprovacaoAlunos.MatricAluno='mineracao_departamentos_ex_aluno'.MatricAluno

SET 'mineracao_departamentos_ex_aluno'.Est=ReprovacaoAlunos.taxa;

UPDATE 'mineracao_departamentos_ex_aluno' INNER JOIN
(SELECT reprovadas.MatricAluno, reprovadas.creditosRepro/total.creditosTot
AS taxa FROM

(SELECT alunos.MatricAluno, SUM(historico.creditos) AS creditosTot FROM
alunos INNER JOIN historico ON alunos.MatricAluno =
historico.MatricAluno JOIN obrigatorias ON obrigatorias.CodDisciplina =
historico.Disciplina WHERE obrigatorias.Departamento='MAT' GROUP BY
alunos.MatricAluno) AS total
INNER JOIN
(SELECT alunos.MatricAluno, SUM(historico.creditos) AS creditosRepro FROM
alunos INNER JOIN historico ON alunos.MatricAluno =
historico.MatricAluno JOIN obrigatorias ON obrigatorias.CodDisciplina =
historico.Disciplina WHERE historico.Aprovacao = 0 AND

```

```

    obrigatorias.Departamento='MAT' GROUP BY alunos.MatricAluno) AS
    reprovadas

ON reprovadas.MatricAluno=total.MatricAluno
GROUP BY reprovadas.MatricAluno) AS ReprovacaoAlunos

ON
    ReprovacaoAlunos.MatricAluno='mineracao_departamentos_ex_aluno'.MatricAluno

SET 'mineracao_departamentos_ex_aluno'.Mat=ReprovacaoAlunos.taxa;

UPDATE 'mineracao_departamentos_ex_aluno' INNER JOIN
(SELECT reprovadas.MatricAluno, reprovadas.creditosRepro/total.creditosTot
    AS taxa FROM

(SELECT alunos.MatricAluno, SUM(historico.creditos) AS creditosTot FROM
    alunos INNER JOIN historico ON alunos.MatricAluno =
    historico.MatricAluno JOIN obrigatorias ON obrigatorias.CodDisciplina =
    historico.Disciplina WHERE obrigatorias.Departamento='IFD' GROUP BY
    alunos.MatricAluno) AS total
INNER JOIN
(SELECT alunos.MatricAluno, SUM(historico.creditos) AS creditosRepro FROM
    alunos INNER JOIN historico ON alunos.MatricAluno =
    historico.MatricAluno JOIN obrigatorias ON obrigatorias.CodDisciplina =
    historico.Disciplina WHERE historico.Aprovacao = 0 AND
    obrigatorias.Departamento='IFD' GROUP BY alunos.MatricAluno) AS
    reprovadas

ON reprovadas.MatricAluno=total.MatricAluno
GROUP BY reprovadas.MatricAluno) AS ReprovacaoAlunos

ON
    ReprovacaoAlunos.MatricAluno='mineracao_departamentos_ex_aluno'.MatricAluno

SET 'mineracao_departamentos_ex_aluno'.IFD=ReprovacaoAlunos.taxa;

UPDATE 'mineracao_departamentos_ex_aluno' INNER JOIN
(SELECT reprovadas.MatricAluno, reprovadas.creditosRepro/total.creditosTot
    AS taxa FROM

(SELECT alunos.MatricAluno, SUM(historico.creditos) AS creditosTot FROM
    alunos INNER JOIN historico ON alunos.MatricAluno =
    historico.MatricAluno JOIN obrigatorias ON obrigatorias.CodDisciplina =
    historico.Disciplina WHERE obrigatorias.Departamento='LIP' GROUP BY
    alunos.MatricAluno) AS total
INNER JOIN
(SELECT alunos.MatricAluno, SUM(historico.creditos) AS creditosRepro FROM
    alunos INNER JOIN historico ON alunos.MatricAluno =
    historico.MatricAluno JOIN obrigatorias ON obrigatorias.CodDisciplina =
    historico.Disciplina WHERE historico.Aprovacao = 0 AND
    obrigatorias.Departamento='LIP' GROUP BY alunos.MatricAluno) AS
    reprovadas

ON reprovadas.MatricAluno=total.MatricAluno

```

```

GROUP BY reprovadas.MatricAluno) AS ReprovacaoAlunos

ON
    ReprovacaoAlunos.MatricAluno='mineracao_departamentos_ex_aluno'.MatricAluno

SET 'mineracao_departamentos_ex_aluno'.Lip=ReprovacaoAlunos.taxa;

UPDATE 'mineracao_departamentos_ex_aluno' INNER JOIN
(SELECT reprovadas.MatricAluno, reprovadas.creditosRepro/total.creditosTot
AS taxa FROM

(SELECT alunos.MatricAluno, SUM(historico.creditos) AS creditosTot FROM
    alunos INNER JOIN historico ON alunos.MatricAluno =
    historico.MatricAluno JOIN obrigatorias ON obrigatorias.CodDisciplina =
    historico.Disciplina WHERE obrigatorias.Departamento='LET' GROUP BY
    alunos.MatricAluno) AS total
INNER JOIN
(SELECT alunos.MatricAluno, SUM(historico.creditos) AS creditosRepro FROM
    alunos INNER JOIN historico ON alunos.MatricAluno =
    historico.MatricAluno JOIN obrigatorias ON obrigatorias.CodDisciplina =
    historico.Disciplina WHERE historico.Aprovacao = 0 AND
    obrigatorias.Departamento='LET' GROUP BY alunos.MatricAluno) AS
    reprovadas

ON reprovadas.MatricAluno=total.MatricAluno
GROUP BY reprovadas.MatricAluno) AS ReprovacaoAlunos

ON
    ReprovacaoAlunos.MatricAluno='mineracao_departamentos_ex_aluno'.MatricAluno

SET 'mineracao_departamentos_ex_aluno'.Let=ReprovacaoAlunos.taxa;

UPDATE 'mineracao_departamentos_ex_aluno' INNER JOIN
(SELECT reprovadas.MatricAluno, reprovadas.creditosRepro/total.creditosTot
AS taxa FROM

(SELECT alunos.MatricAluno, SUM(historico.creditos) AS creditosTot FROM
    alunos INNER JOIN historico ON alunos.MatricAluno =
    historico.MatricAluno JOIN obrigatorias ON obrigatorias.CodDisciplina =
    historico.Disciplina WHERE obrigatorias.Departamento='Fil' GROUP BY
    alunos.MatricAluno) AS total
INNER JOIN
(SELECT alunos.MatricAluno, SUM(historico.creditos) AS creditosRepro FROM
    alunos INNER JOIN historico ON alunos.MatricAluno =
    historico.MatricAluno JOIN obrigatorias ON obrigatorias.CodDisciplina =
    historico.Disciplina WHERE historico.Aprovacao = 0 AND
    obrigatorias.Departamento='Fil' GROUP BY alunos.MatricAluno) AS
    reprovadas

ON reprovadas.MatricAluno=total.MatricAluno
GROUP BY reprovadas.MatricAluno) AS ReprovacaoAlunos

ON
    ReprovacaoAlunos.MatricAluno='mineracao_departamentos_ex_aluno'.MatricAluno

```



```
SET 'mineracao_departamentos_ex_aluno'.Fil=ReprovacaoAlunos.taxa;
```

```
UPDATE 'mineracao_departamentos_ex_aluno' SET 'Cic'=  
CASE WHEN 'Cic' = 0 THEN '10'  
    WHEN 'Cic'>0 AND 'Cic'<0.1 THEN '11'  
    WHEN 'Cic'>=0.1 AND 'Cic'<0.2 THEN '12'  
    WHEN 'Cic'>=0.2 AND 'Cic'<0.3 THEN '13'  
    WHEN 'Cic'>=0.3 AND 'Cic'<0.4 THEN '14'  
    WHEN 'Cic'>=0.4 AND 'Cic'<0.5 THEN '15'  
    WHEN 'Cic'>=0.5 AND 'Cic'<0.6 THEN '16'  
    WHEN 'Cic'>=0.6 AND 'Cic'<0.7 THEN '17'  
    WHEN 'Cic'>=0.7 AND 'Cic'<0.8 THEN '18'  
    WHEN 'Cic'>=0.8 AND 'Cic'<0.9 THEN '19'  
    WHEN 'Cic'>=0.9 AND 'Cic'<1 THEN '20'  
    WHEN 'Cic'=1 THEN '21'  
end;
```

```
UPDATE 'mineracao_departamentos_ex_aluno' SET 'Est'=  
CASE WHEN 'Est' = 0 THEN '10'  
    WHEN 'Est'>0 AND 'Est'<0.1 THEN '11'  
    WHEN 'Est'>=0.1 AND 'Est'<0.2 THEN '12'  
    WHEN 'Est'>=0.2 AND 'Est'<0.3 THEN '13'  
    WHEN 'Est'>=0.3 AND 'Est'<0.4 THEN '14'  
    WHEN 'Est'>=0.4 AND 'Est'<0.5 THEN '15'  
    WHEN 'Est'>=0.5 AND 'Est'<0.6 THEN '16'  
    WHEN 'Est'>=0.6 AND 'Est'<0.7 THEN '17'  
    WHEN 'Est'>=0.7 AND 'Est'<0.8 THEN '18'  
    WHEN 'Est'>=0.8 AND 'Est'<0.9 THEN '19'  
    WHEN 'Est'>=0.9 AND 'Est'<1 THEN '20'  
    WHEN 'Est'=1 THEN '21'  
end;
```

```
UPDATE 'mineracao_departamentos_ex_aluno' SET 'Mat'=  
CASE WHEN 'Mat' = 0 THEN '10'  
    WHEN 'Mat'>0 AND 'Mat'<0.1 THEN '11'  
    WHEN 'Mat'>=0.1 AND 'Mat'<0.2 THEN '12'  
    WHEN 'Mat'>=0.2 AND 'Mat'<0.3 THEN '13'  
    WHEN 'Mat'>=0.3 AND 'Mat'<0.4 THEN '14'  
    WHEN 'Mat'>=0.4 AND 'Mat'<0.5 THEN '15'  
    WHEN 'Mat'>=0.5 AND 'Mat'<0.6 THEN '16'  
    WHEN 'Mat'>=0.6 AND 'Mat'<0.7 THEN '17'  
    WHEN 'Mat'>=0.7 AND 'Mat'<0.8 THEN '18'  
    WHEN 'Mat'>=0.8 AND 'Mat'<0.9 THEN '19'  
    WHEN 'Mat'>=0.9 AND 'Mat'<1 THEN '20'  
    WHEN 'Mat'=1 THEN '21'  
end;
```

```
UPDATE 'mineracao_departamentos_ex_aluno' SET 'IFD'=  
CASE WHEN 'IFD' = 0 THEN '10'  
    WHEN 'IFD'>0 AND 'IFD'<0.1 THEN '11'  
    WHEN 'IFD'>=0.1 AND 'IFD'<0.2 THEN '12'  
    WHEN 'IFD'>=0.2 AND 'IFD'<0.3 THEN '13'  
    WHEN 'IFD'>=0.3 AND 'IFD'<0.4 THEN '14'  
    WHEN 'IFD'>=0.4 AND 'IFD'<0.5 THEN '15'
```

```

WHEN 'IFD'>=0.5 AND 'IFD'<0.6 THEN'16'
WHEN 'IFD'>=0.6 AND 'IFD'<0.7 THEN'17'
WHEN 'IFD'>=0.7 AND 'IFD'<0.8 THEN'18'
WHEN 'IFD'>=0.8 AND 'IFD'<0.9 THEN'19'
WHEN 'IFD'>=0.9 AND 'IFD'<1 THEN'20'
WHEN 'IFD'=1 THEN'21'
end;

UPDATE 'mineracao_departamentos_ex_aluno' SET 'Lip'='
CASE WHEN 'Lip' = 0 THEN '10'
    WHEN 'Lip'>0 AND 'Lip'<0.1 THEN'11'
    WHEN 'Lip'>=0.1 AND 'Lip'<0.2 THEN'12'
    WHEN 'Lip'>=0.2 AND 'Lip'<0.3 THEN'13'
    WHEN 'Lip'>=0.3 AND 'Lip'<0.4 THEN'14'
    WHEN 'Lip'>=0.4 AND 'Lip'<0.5 THEN'15'
    WHEN 'Lip'>=0.5 AND 'Lip'<0.6 THEN'16'
    WHEN 'Lip'>=0.6 AND 'Lip'<0.7 THEN'17'
    WHEN 'Lip'>=0.7 AND 'Lip'<0.8 THEN'18'
    WHEN 'Lip'>=0.8 AND 'Lip'<0.9 THEN'19'
    WHEN 'Lip'>=0.9 AND 'Lip'<1 THEN'20'
    WHEN 'Lip'=1 THEN'21'
end;

UPDATE 'mineracao_departamentos_ex_aluno' SET 'Let'='
CASE WHEN 'Let' = 0 THEN '10'
    WHEN 'Let'>0 AND 'Let'<0.1 THEN'11'
    WHEN 'Let'>=0.1 AND 'Let'<0.2 THEN'12'
    WHEN 'Let'>=0.2 AND 'Let'<0.3 THEN'13'
    WHEN 'Let'>=0.3 AND 'Let'<0.4 THEN'14'
    WHEN 'Let'>=0.4 AND 'Let'<0.5 THEN'15'
    WHEN 'Let'>=0.5 AND 'Let'<0.6 THEN'16'
    WHEN 'Let'>=0.6 AND 'Let'<0.7 THEN'17'
    WHEN 'Let'>=0.7 AND 'Let'<0.8 THEN'18'
    WHEN 'Let'>=0.8 AND 'Let'<0.9 THEN'19'
    WHEN 'Let'>=0.9 AND 'Let'<1 THEN'20'
    WHEN 'Let'=1 THEN'21'
end;

UPDATE 'mineracao_departamentos_ex_aluno' SET 'Fil'='
CASE WHEN 'Fil' = 0 THEN '10'
    WHEN 'Fil'>0 AND 'Fil'<0.1 THEN'11'
    WHEN 'Fil'>=0.1 AND 'Fil'<0.2 THEN'12'
    WHEN 'Fil'>=0.2 AND 'Fil'<0.3 THEN'13'
    WHEN 'Fil'>=0.3 AND 'Fil'<0.4 THEN'14'
    WHEN 'Fil'>=0.4 AND 'Fil'<0.5 THEN'15'
    WHEN 'Fil'>=0.5 AND 'Fil'<0.6 THEN'16'
    WHEN 'Fil'>=0.6 AND 'Fil'<0.7 THEN'17'
    WHEN 'Fil'>=0.7 AND 'Fil'<0.8 THEN'18'
    WHEN 'Fil'>=0.8 AND 'Fil'<0.9 THEN'19'
    WHEN 'Fil'>=0.9 AND 'Fil'<1 THEN'20'
    WHEN 'Fil'=1 THEN'21'
end;

ALTER TABLE 'mineracao_departamentos_ex_aluno'
DROP 'MatricAluno';

```